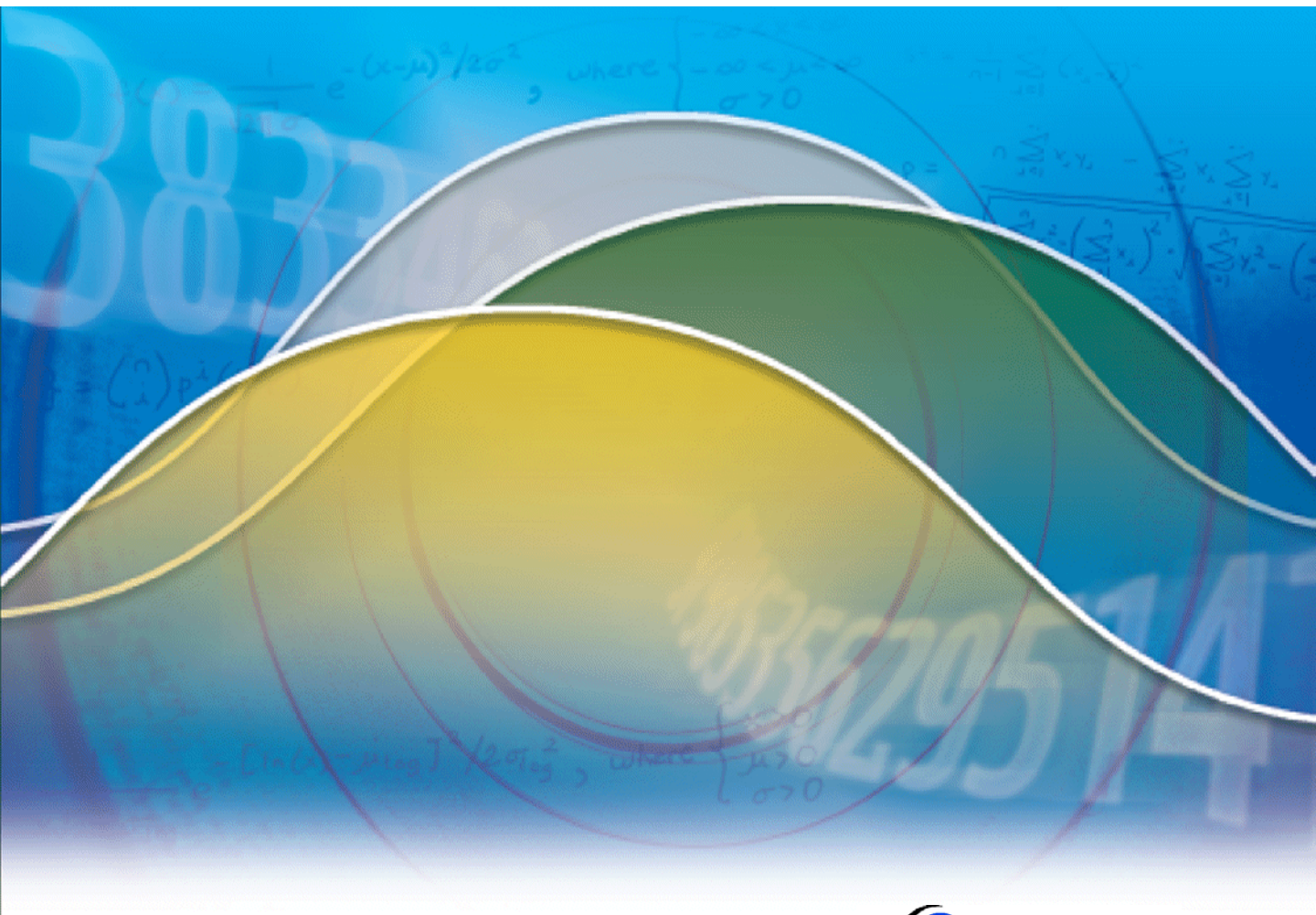


CB Predictor™ 1.6

User Manual



This manual, and the software described in it, are furnished under license and may only be used or copied in accordance with the terms of the license agreement. Information in this document is provided for informational purposes only, is subject to change without notice, and does not represent a commitment as to merchantability or fitness for a particular purpose by Decisioneering, Inc.

No part of this manual may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording, for any purpose without the express written permission of Decisioneering, Inc.

Written, designed, and published in the United States of America.

To purchase additional copies of this document, contact the Technical Services or Sales Department at the address below:

Decisioneering, Inc.
1515 Arapahoe St., Suite 1311
Denver, Colorado, USA 80202

Phone: +1 303-534-1515 Toll-free sales: 1-800-289-2550 Fax: 1-303-534-4818

© 1988-2006, Decisioneering, Inc.

Decisioneering® is a registered trademark of Decisioneering, Inc.


Crystal Ball® is a registered trademark of Decisioneering, Inc.


CB Predictor™ is a trademark of Decisioneering, Inc.

OptQuest® is a registered trademark of Optimization Technologies, Inc.

Microsoft® is a registered trademark of Microsoft Corporation in the U.S. and other countries.

FLEXlm™ is a trademark of Macrovision Corporation.

 Chart FX® is a registered trademark of Software FX, Inc.

 is a registered trademark of Frontline Systems, Inc.

Other product names mentioned herein may be trademarks and/or registered trademarks of the respective holders.

Contents

Welcome to CB Predictor™

Who should use this program?	1
How this manual is organized	2
Additional resources	3
Technical support	3
Training	3
Consulting	4
Conventions used in this manual	4
Screen capture notes	5
Example file names	5

Chapter 1: Getting Started with CB Predictor

What is time-series forecasting?	8
Shampoo Sales tutorial	9
How CB Predictor works	16
Toledo Gas tutorial	17
Using regression	19
Selecting results	21
Previewing results	24
Summary	26

Chapter 2: Understanding the Terminology

Forecasting	34
Time-series forecasting	35
Nonseasonal smoothing methods	36
Single moving average	36
Double moving average	36
Single exponential smoothing (SES)	37
Holt's double exponential smoothing (DES)	37
Nonseasonal smoothing parameters	38
Seasonal smoothing methods	39
Seasonal Additive Smoothing	39
Seasonal Multiplicative Smoothing	39
Holt-Winters' Additive Seasonal Smoothing	40
Holt-Winters' Multiplicative Seasonal Smoothing	41
Seasonal smoothing parameters	41
Time-series forecasting error measures	42
RMSE	42
MAD	43
MAPE	43
Time-series forecasting techniques	43
Standard forecasting	43

Contents

Simple lead forecasting	44
Weighted lead forecasting	44
Holdout	45
Time-series forecasting statistics.....	46
Theil's U	46
Multiple linear regression.....	46
Regression methods.....	48
Standard regression	48
Forward stepwise regression	48
Iterative stepwise regression	49
Regression statistics	50
R2	50
Adjusted R2	50
SSE	50
F statistic	51
t statistic	51
p	51
Durbin-Watson	51
Historical data statistics	52
mean	52
standard deviation	52
minimum	53
maximum	53
Ljung-Box statistic	53

Chapter 3: Forecasting with CB Predictor

Overview.....	56
Creating a spreadsheet with historical data.....	56
Loading and starting CB Predictor	57
Guidelines for using CB Predictor.....	58
Selecting historical data	59
Automatic data selection	60
Manual data selection before starting CB Predictor	60
Manual data selection within CB Predictor	60
Specifying data arrangement	61
Viewing your historical data	61
Identifying time periods and seasonality	63
Using multiple linear regression	64
Selecting time-series forecasting methods	66
Selecting error measures	69
Selecting forecasting techniques	70
Entering the number of periods to forecast	71
Selecting a confidence interval	71

Select your results	72
Previewing and running the forecast	73
Analyzing the results	74
Pasted forecasts	74
Charts	75
Reports	76
Results table	76
Methods table	77
Customizing reports and charts	79

Chapter 4: Examples

Overview	82
Inventory control.....	82
Company finances	86
Human resources	88

Chapter 5: Using CB Predictor with Crystal Ball

Using CB Predictor with Crystal Ball.....	94
Automatically generating Crystal Ball assumptions	94
Time-series forecast method results	95
Multiple linear regression results	95
Monica's Bakery example revisited.....	96

Appendix A: The CB Predictor Wizard

Input Data tab	100
View Historical Data dialog	102
View Historical Data — Autocorrelations view	103
Data Attributes tab.....	104
Regression Variables dialog	106
Stepwise Options dialog	108
Method Gallery tab.....	110
Parameters dialogs	111
Advanced Options dialog	113
Results tab	115
Preferences dialog	117
Preferences dialog, Paste tab	118
Preferences dialog, Report tab	119
Preferences dialog, Chart tab	121
Preferences dialog, Results Table tab	122
Preferences dialog, Methods Table tab	123
Preview Forecast dialog	125

Contents

Appendix B: Time-Series Forecasting Method Formulas

Nonseasonal forecasting method formulas	128
Double moving average	128
Single exponential smoothing	128
Holt's double exponential smoothing	128
Seasonal forecasting method formulas	129
Seasonal additive smoothing	129
Seasonal multiplicative smoothing	129
Holt-Winters' additive seasonal smoothing	130
Holt-Winters' multiplicative seasonal smoothing	131

Appendix C: Error Measure and Statistic Formulas

Time-series forecast error measures	134
RMSE	134
MAD	135
MAPE	135
Confidence intervals	136
Time-series forecast statistics	137
Theil's U statistic	137
Autocorrelation statistics	138
Durbin-Watson	138
Autocorrelation	139
Ljung-Box statistic	139

Appendix D: Techniques for Finding Regression Coefficients

Finding regression coefficients	142
Least squares	142
Singular value decomposition	143

Appendix E: Regression Statistic Formulas

Regression statistics	146
R ²	146
Adjusted R ²	146
SSE	147
F	147
t	149
p	150

Appendix F: Error Messages

Error messages	152
----------------------	-----

Bibliography	159
---------------------------	-----

Glossary	161
-----------------------	-----

Index	169
--------------------	-----

Welcome to CB Predictor™

Welcome to CB Predictor, a powerful addition to the Crystal Ball suite of decision intelligence products!

Forecasting is an important part of many business decisions. Every organization needs to set goals, try to predict future events, and then act to fulfill the goals. As the timeliness of market actions becomes more important, the need for accurate planning and forecasting throughout an organization is essential to get ahead. The difference between good and bad forecasting can affect the success of an entire organization.

CB Predictor is an easy-to-use, graphically oriented forecasting add-in for Microsoft Excel spreadsheet users. If you have historical data in your spreadsheet, CB Predictor analyzes your data for trends and seasonal variations. It then predicts future values based on this information. You can answer questions such as, “What are the likely figures for next quarter’s sales?” or, “How much material do we need to have on hand?” With CB Predictor, you no longer need to pull these numbers out of thin air. Instead, you can rely on robust, statistically proven techniques to create these predictions accurately.

To get started, all you need is a spreadsheet containing historical data. From there, this manual guides you step by step, explaining forecasting terms and results.

Who should use this program?

CB Predictor is for the planner and forecaster in every organization, from the sales manager predicting next quarter’s sales, to the marketing specialist forecasting the results of an advertising campaign, to the financial officer projecting likely revenue figures and cash flows. CB Predictor has been developed with a wide range of forecasting applications in mind.

You don't need highly advanced statistical or computer knowledge to use CB Predictor to its full potential. All you need is a basic working knowledge of your computer and Microsoft Excel, plus a solid understanding of your own business fundamentals.

As an added benefit, you can automatically save CB Predictor forecasts as Crystal Ball assumptions for immediate use in powerful risk analysis models. See Chapter 5 for more information.

CB Predictor runs on several versions of Microsoft Windows and Microsoft Excel. For a list of required hardware and software, see README.htm in the main Crystal Ball installation folder (by default, C:\Program Files\Decisioneering\Crystal Ball 7).

How this manual is organized

The manual includes the following:

- **Chapter 1 — “Getting Started with CB Predictor”**

This chapter contains two tutorials designed to give you a quick overview of CB Predictor’s features and to show you how to use it. Read this chapter if you need a basic understanding of CB Predictor.

- **Chapter 2 — “Understanding the Terminology”**

This chapter contains a thorough description of forecasting and statistical terms. Read this chapter carefully if your statistics background is limited or if you need a review of how terms are used in CB Predictor.

- **Chapter 3 — “Forecasting with CB Predictor”**

This chapter contains step-by-step procedures for using all the features in CB Predictor.

- **Chapter 4 — “Examples”**

This chapter contains example data from various fields.

- **Chapter 5 — “Using CB Predictor with Crystal Ball”**

This chapter contains descriptions of how to use CB Predictor with Crystal Ball. It also has some examples that are Crystal Ball-specific.

- **Appendices**

- A — “The CB Predictor Wizard”

Settings for each tab of the CB Predictor wizard dialog.

- B — “Time-Series Forecasting Method Formulas”

The formulas used for the time-series forecasting methods.

- C — “Error Measure and Statistic Formulas”

The formulas used for time-series error measures and other statistics.

- D — “Techniques for Finding Regression Coefficients”

How the program determines the coefficients of regression equations.

- E — “Regression Statistic Formulas”

The formulas used to calculate the statistics to gauge the quality of the regression and the coefficients.

- F — “Error Messages”

A list of the most important CB Predictor error messages with directions for error recovery.

- **Bibliography**

A list of related publications, including statistics textbooks.

- **Glossary**

A compilation of terms specific to CB Predictor as well as statistical terms used in this manual.

- **Index**

An alphabetical list of subjects and corresponding page numbers.

Additional resources

Decisioneering, Inc. offers these additional resources to help you use our products most effectively.

Technical support

Technical Support is available for all registered customers with a current maintenance agreement and a valid license authorization code. There are a number of ways to reach Technical Support described in the README file in the Crystal Ball installation folder. Online, see:

<http://support.crystalball.com>

Training

Decisioneering’s Training group offers a variety of courses throughout the year to help improve how you make decisions. For more information about Decisioneering courses, call one of these numbers Monday through Friday, between 8:00 a.m. and 5:00 p.m. Mountain Time: 1-800-289-2550 (toll free in US) or +1 303-534-1515, or visit the Decisioneering Web site:

<http://www.crystalball.com/training>

Consulting

Decisioneering's Services group provides consulting services including the full range of risk analysis techniques from simulation, optimization, advanced statistical analysis and exact probability calculations, to strategic thinking, training, expert elicitation, and results communication to management. To learn more about these consulting services, call 1-800-289-2550 Monday through Friday, between 8:00 A.M. and 5:00 P.M. Mountain Time or see our Web site at:

<http://www.crystalball.com/consulting>

Conventions used in this manual

This manual uses the following conventions:

- Text separated by > symbols means you select menu commands in the sequence shown, starting from the left. The following example means that you select Exit from the File menu:

1. Select File > Exit.

- Steps with attached icons mean you can click the icon instead of manually selecting the menu commands in the text.
- Notes provide additional information, expanding on the text. There are five categories of notes:

CB Predictor Note: Notes that provide additional directions or information about using CB Predictor.

Crystal Ball Note: Notes that provide additional directions or information about using Crystal Ball.

Statistical Note: Notes that provide additional information about statistics.

Excel Note: Notes that provide additional information about using the program with Microsoft Excel.

Windows Note: Notes that provide additional information about using the program on a Windows system.

Screen capture notes

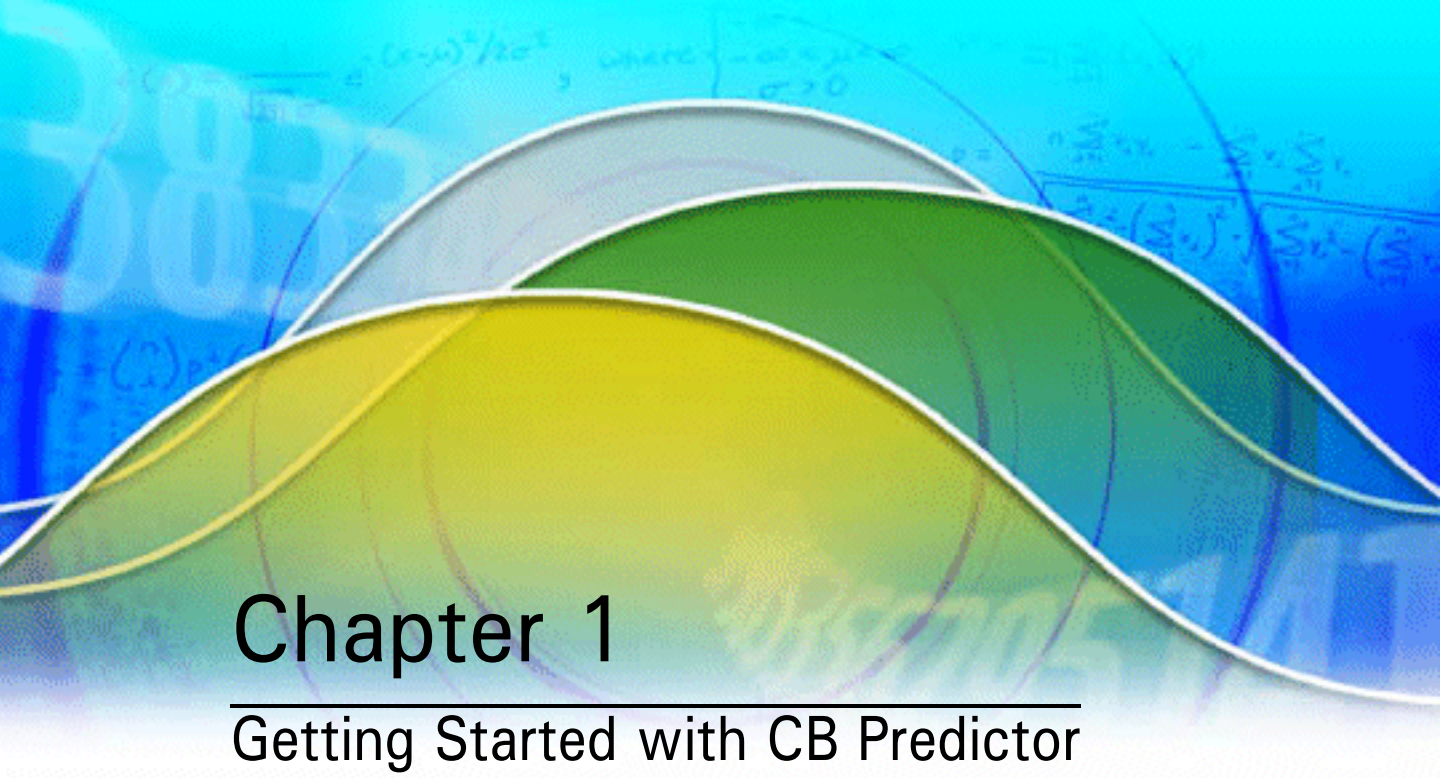
The screen captures in this manual were taken in Excel 2003 for Windows XP Professional.

Due to round-off differences between various system configurations, you might notice slightly different calculated results than those shown in the examples.

Example file names

Example names are listed in full wherever given. You can find the example files in their Examples folder; by default C:\Program Files\Decisioneering\Crystal Ball 7\Examples\CB Predictor Examples. For example, to find the “Toledo Gas” example, select “Toledo Gas.xls”.

Welcome



Chapter 1

Getting Started with CB Predictor

In this chapter

- What is time-series forecasting?
- Shampoo Sales tutorial
- How CB Predictor works
- Toledo Gas tutorial

This chapter has two tutorials, a short one and a longer one, that provide an overview of CB Predictor's features. The first tutorial, Shampoo Sales, is ready to run, using most of the defaults. The results forecast the next quarter's worth of shampoo sales.

The second tutorial, Toledo Gas, uses regression, previews the results, and customizes some features. The results predict the next year's worth of residential gas usage for Toledo.

Now spend some time learning how CB Predictor can help you forecast the future.

What is time-series forecasting?

“Prediction is very difficult, especially if it’s about the future.”

- Niels Bohr

Nobel laureate in Physics

How often are you asked to estimate future sales? Staffing requirements? Budget items? Do you need to predict numbers for the next year using historical numbers that are higher in the summer than the winter? Or numbers that rise and fall with the Dow Jones industrial average? If so, CB Predictor can help.

Glossary Term:

time series—

A set of values that are ordered in equally spaced intervals of time.

Glossary Term:

level—

A starting point for the forecast.

Glossary Term:

trend—

A long-term increase or decrease in time-series data.

Glossary Term:

seasonality—

The change that seasonal factors cause in a data series. For example, if sales increase during the Christmas season and during the summer, the data is seasonal with a one-year period.

Glossary Term:

error—

The difference between the actual data values and the forecasted data values.

If you have historical *time-series* data, you can use CB Predictor to examine your historical data and predict future trends. You no longer have to assume that sales will be same as last quarter or that you will spend 10% more next year on expenses than this year. CB Predictor applies a battery of sophisticated statistical methods to your data series to find results that meet the strictest confidence requirements.

CB Predictor uses two types of forecasting: time-series forecasting and multiple linear regression. Time-series forecasting breaks down your historical data into four components: *level*, *trend*, *seasonality*, and *error*. CB Predictor analyzes each of these components and then projects them into the future to predict likely results.

When you know that outside influences have an effect on the variable you want to forecast, you need regression. Regression takes historical data from the influencing variables and determines the mathematical relationship between these variables and your target variable. It then uses time-series forecasting methods to forecast the influencing variables and combines the results mathematically to forecast your target variable.

After finding the best forecast for your data, CB Predictor pastes the forecasted values into your spreadsheet. You can also request detailed output that includes statistics, charts, reports, and PivotTables. CB Predictor can also create your forecasted values as Crystal Ball assumptions, ready for a “what-if” simulation.

Shampoo Sales tutorial

The easiest way to understand what CB Predictor does is to apply it to a simple example. In this example, you are sales manager for Tropical Cosmetics Co. The company's latest product, shampoo with tropical ingredients, has been in the marketplace for almost a year now. The vice president of marketing wants you to forecast the rest of the year's sales of shampoo and decide whether to recommend investing in advertising or enhancements for this product.

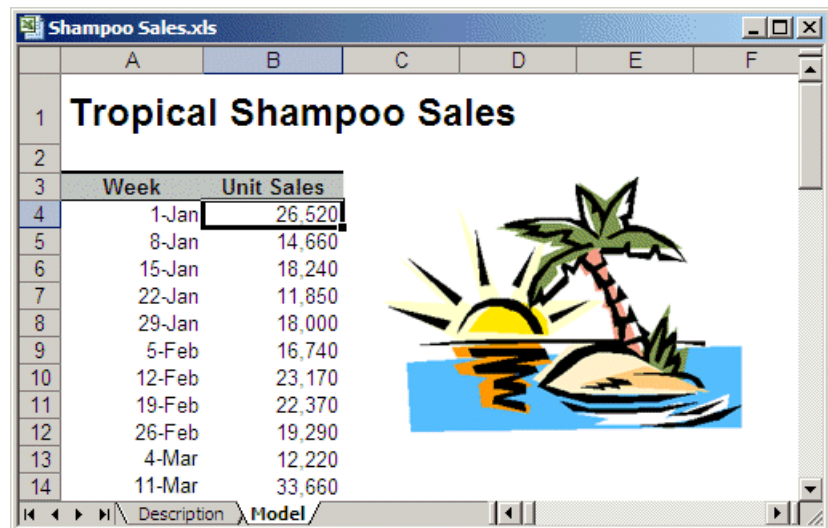
You have the weekly sales numbers for the last nine months.

To begin the tutorial:

1. **Start Crystal Ball and Excel.**
2. **Open the Shampoo Sales spreadsheet from the Examples folder.**

By default, the file is stored in this folder:

C:\Program Files\Decisioneering\Crystal Ball 7\Examples\CB Predictor Examples.



Week	Unit Sales
1-Jan	26,520
8-Jan	14,660
15-Jan	18,240
22-Jan	11,850
29-Jan	18,000
5-Feb	16,740
12-Feb	23,170
19-Feb	22,370
26-Feb	19,290
4-Mar	12,220
11-Mar	33,660

Figure 1.1 Shampoo Sales spreadsheet

In this spreadsheet, there is one column of Tropical shampoo sales data next to a column with dates from January 1, 2004 until September 23, 2004. You need to forecast sales through the end of the year, December 31, 2004.

3. Select cell B4.

Selecting any one cell in your data range, headers, or date range initiates CB Predictor's "Intelligent Input" to select all the filled, adjacent cells.

4. Select Run > CB Predictor.

Note: This command is only available if no simulation is running and the last run was reset. If necessary, wait for a simulation to stop or reset the last simulation.

The CB Predictor dialog, or wizard, opens to the Input Data tab as shown in Figure 1.2.

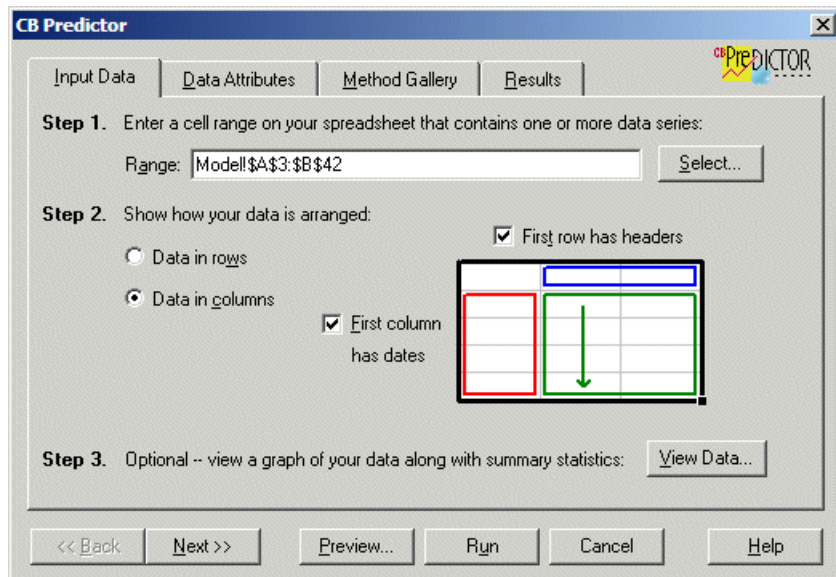


Figure 1.2 CB Predictor wizard, Input Data tab

When you select any one cell in the data range before you start the wizard, CB Predictor's Intelligent Input guesses:

- Your data series (in this case, A3:B42)
- Whether your data are in columns or rows
- Whether you have headers at the beginning of your data
- Whether your first column or row contains dates or time periods

5. Click Next.

The Data Attributes tab appears as shown in Figure 1.3.

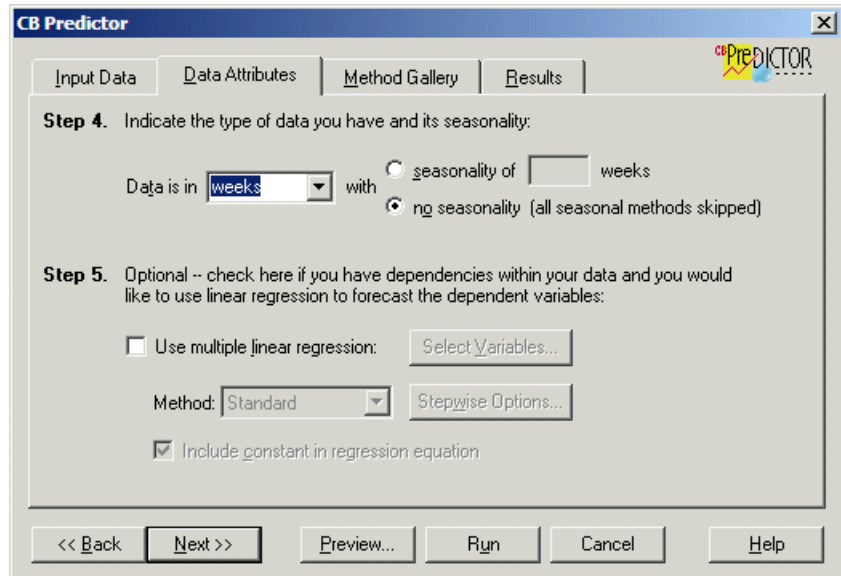


Figure 1.3 CB Predictor wizard, Data Attributes tab

6. Under Step 4:

- a. Select “weeks” from the Data Is In list.
- b. Set the data to have no seasonality.

You have less than two complete seasons (cycles) of data, so cannot use seasonality.

7. Under Step 5, make sure that Use Multiple Linear Regression is *not* checked.

You did not choose regression because you have only one series of data, so there are no dependencies between series requiring regression.

8. Click Next.

The Method Gallery tab appears as shown in Figure 1.4.

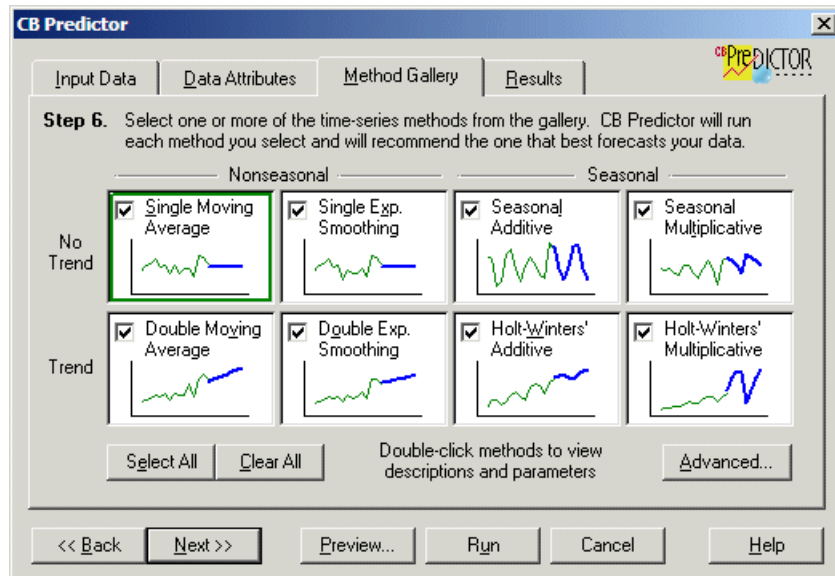


Figure 1.4 CB Predictor wizard, Method Gallery tab

9. Click Select All.

This selects all the time-series forecasting methods, but CB Predictor doesn't use the seasonal methods, since you indicated that your data were not seasonal.

CB Predictor forecasts your values using each of the selected methods and ranks them according to how well they fit the historical data. CB Predictor uses the seasonal methods as well as the nonseasonal methods if you indicate on the Data Attributes tab that your data series have seasonality.

10. Click Next.

The Results tab appears as shown in Figure 1.5. The only output selected by default is Paste Forecast, which adds the forecasted values to the end of your historical data as shown in Figure 1.5.

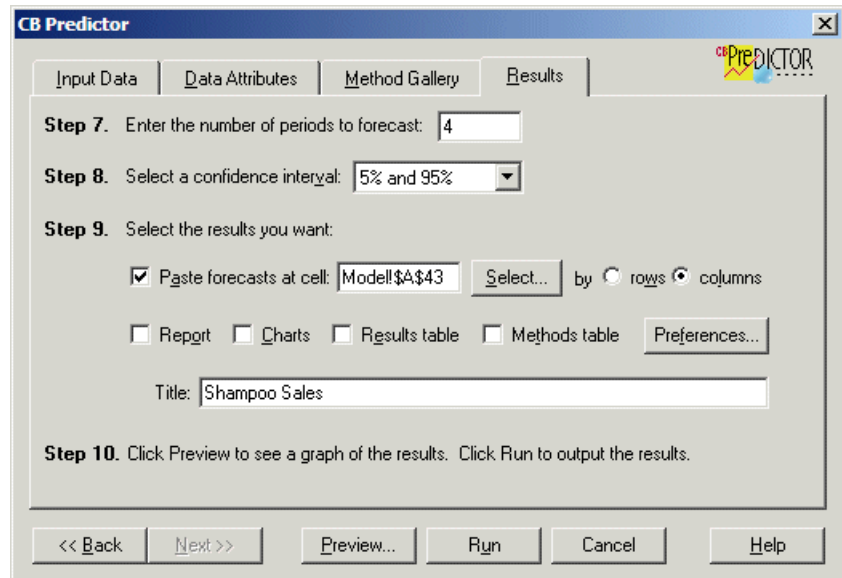


Figure 1.5 CB Predictor wizard, Results tab

11. Under Step 7, forecast the weekly sales for the rest of the year by entering 13 in the field.

12. Click Preview.

The Preview Forecast dialog appears. It presents a graph with historical data, fitted data, forecast values, and confidence intervals as shown in Figure 1.6.

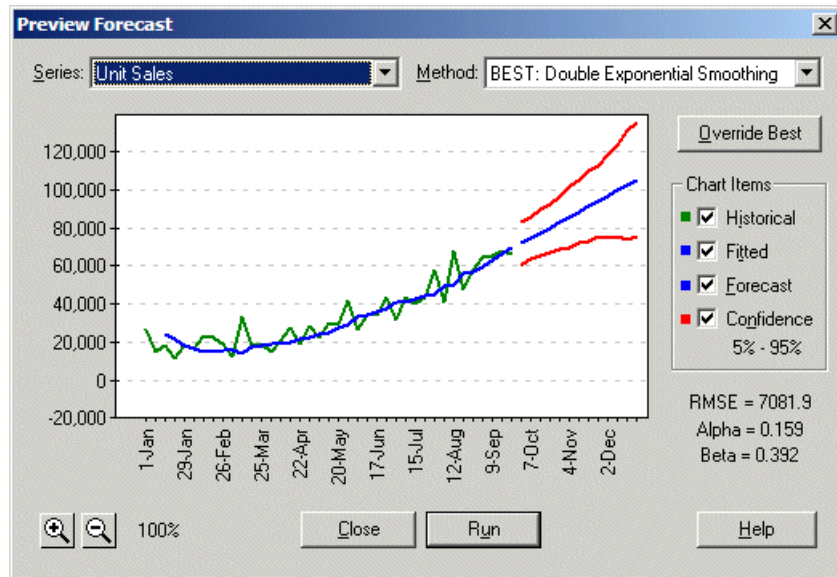


Figure 1.6 Preview Forecast dialog

13. In the Preview Forecast dialog, click the Method field.

The field lists all the methods CB Predictor tried, in order from the best-fitting method (designated by the word “Best”) to the worst-fitting method. CB Predictor calculates the forecasted values from the method that best fits the historical data. In this case, the method is Double Exponential Smoothing.

The forecasted values appear as a blue line extending to the right of the historical data (green) and the fitted values (also in blue). Above and below the forecasted values is the confidence interval (in red), showing the 5th and 95th percentiles of the forecasted values.

14. Click Run.

The program pastes the forecasted values at the end of the historical data (in bold), extending the date series as well. The forecasted values were forecasted using the best method, as shown in the Preview dialog.

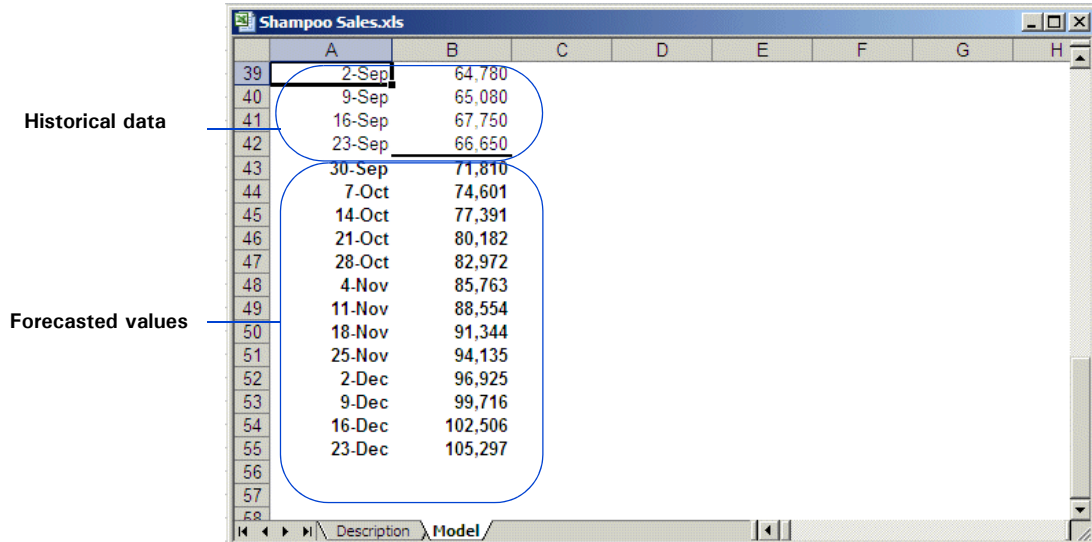


Figure 1.7 Pasted shampoo sales values

Based on the results, you complete your memo to upper management. Current strategies seem to be working so you recommend funding another project instead.

INTEROFFICE MEMO

DATE: 11/11/2004

TO: MARKETING VP

FROM: SALES MANAGER

RE: SALES OF NEW TROPICAL SHAMPOO

Per your request, I have painstakingly analyzed the Tropical Shampoo sales data for the year-to-date. The analysis indicates that the sales trend, with the current marketing strategy, is sufficient to drive the sales to the projected product goal of 100,000 units per week by the end of the year.

With this in mind, I'd like to recommend the re-direction of the marketing dollars to the new release our hair conditioner - SunOut!

Figure 1.8 Analysis memo

How CB Predictor works

Most historical or time-based data contain some kind of underlying trend or seasonal pattern. However, most historical data also contain random fluctuations — “noise”— that make it difficult to detect these trends and patterns without using a computer. CB Predictor uses sophisticated time-series methods to analyze the underlying structure of your data. It then projects the trends and patterns to predict future values.

When you run CB Predictor, it tries each time-series method on your data and calculates a mathematical measure of goodness-of-fit. CB Predictor selects the method with the best goodness-of-fit as the method that will yield the most accurate forecast. CB Predictor performs this selection automatically for you, but you can also select individual methods manually or override the method CB Predictor recommends with a different one.

The final forecast shows the most likely continuation of your data. Keep in mind that all these methods assume that some aspects of the historical trend or pattern will continue into the future. However, the farther out you forecast, the higher the likelihood that events will diverge from past behavior, and the less confident you can be of the results. To help you gauge the reliability of your forecast, CB Predictor provides a confidence interval indicating the degree of uncertainty around your forecast.

The final step of forecasting involves interpreting the results and integrating them into your decision-making process. CB Predictor provides comprehensive output to assist you in this process. You can request detailed reports, customizable Excel charts and graphs, and summary PivotTables. CB Predictor can also automatically create Crystal Ball assumptions for each forecasted value, letting you integrate risk analysis into the process.

Toledo Gas tutorial

Suppose you work for Toledo Gas Company in the Residential Division. The Public Utilities Commission requires you to predict the gas usage for the coming year to make sure that the company can meet the demand.

To start the tutorial:

1. **In Excel with Crystal Ball loaded, open the Toledo Gas spreadsheet, Toledo Gas.xls.**

To find this example, select Help > Crystal Ball > Crystal Ball Examples and locate it in the list of CB Predictor examples, or browse to its folder. By default, the file is stored in this folder:

C:\Program Files\Decisioneering\Crystal Ball 7\Examples\CB Predictor Examples.

When you double-click the link or the file, the Toledo Gas spreadsheet appears in Excel.

Date	Usage (ft³)	Occupancy Permits	Average Temperature (Degrees F)	Cost of Natural Gas per ccf (Dollars)
Oct-95	306.00	406	24.46	4.82
Nov-95	266.00	472	24.86	5.21
Dec-95	222.00	599	35.68	5.82
Jan-96	179.00	714	46.54	6.78
Feb-96	104.00	892	59.27	6.41
Mar-96	81.00	874	67.45	10.18
Apr-96	48.00	896	72.41	9.12
May-96	53.00	890	71.26	8.83
Jun-96	53.00	661	61.04	8.69
Jul-96	74.00	371	50.34	9.60

Figure 1.9 Toledo Gas spreadsheet

2. **Select cell C5.**
3. **Choose Run > CB Predictor.**

The Input Data tab appears. CB Predictor's Intelligent Input feature selects all the data from cell B4 to cell F100.

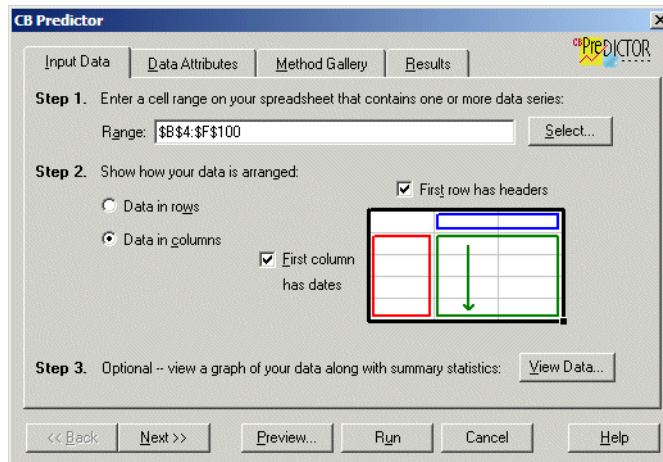


Figure 1.10 Input Data tab for Toledo Gas model

4. Click Next.

The Data Attributes tab appears.

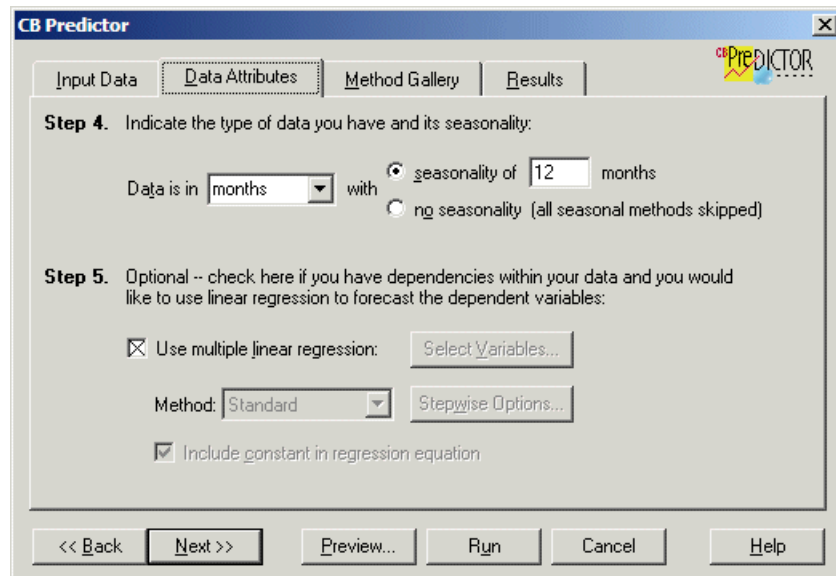


Figure 1.11 Data Attributes tab

5. Confirm that in Step 4, settings are “Data is in months with seasonality of 12 months.”

Using regression

Through research, you know that your residential gas usage is primarily affected by three variables: new home starts, the temperature, and the price of natural gas. However, you aren't sure how much effect each has on gas usage.

Because you have independent variables affecting a dependent variable (the variable that you are interested in), this forecast requires *regression*. For regression, CB Predictor uses a technique called HyperCasting™, which in one easy step:

Glossary Term:
regression—
A process that models a
dependent variable as a
function of other
explanatory variables.

- a. **Creates an equation that defines the mathematical relationship between the independent variables and your dependent variable.**
- b. **Forecasts each independent variable using time-series forecasting methods.**
- c. **Uses the equation it created in the first step, combining the forecasted independent variable values, to create the forecast for the dependent variable.**

In the Toledo Gas spreadsheet, the dependent variable is the historical residential gas usage. The independent variables are:

- Number of occupancy permits issued (new housing completions)
- Average temperature per month
- Unit cost of natural gas

To resume your tutorial:

6. Under Step 5, select Use Multiple Linear Regression.

The Regression Variables dialog appears.

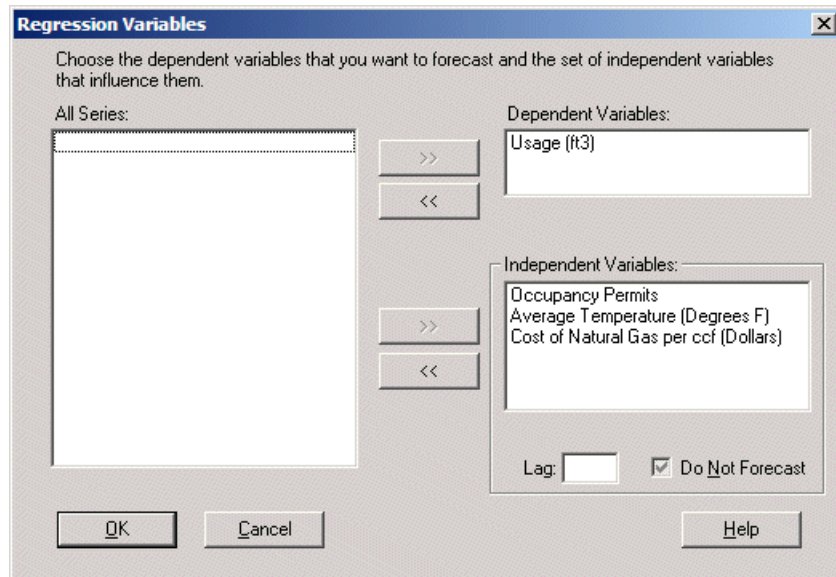


Figure 1.12 Regression Variables dialog

7. Confirm that Usage appears in the Dependent Variables list.

If it does not:

- Select Usage in the All Series list.
- Click >> next to the Dependent Variables list.

The Usage variable moves to the Dependent Variables list.

8. Confirm that Occupancy Permits, Average Temperature, and Cost Of Nature Gas Per Ccf appear in the Independent Variables list.

If they do not:

- Select Occupancy Permits, Average Temperature, and Cost Of Natural Gas Per Ccf in the All Series list.
- Click >> next to the Independent Variables list.

The variables move to the Independent Variables list.

9. Click OK.

The Data Attributes tab reappears.

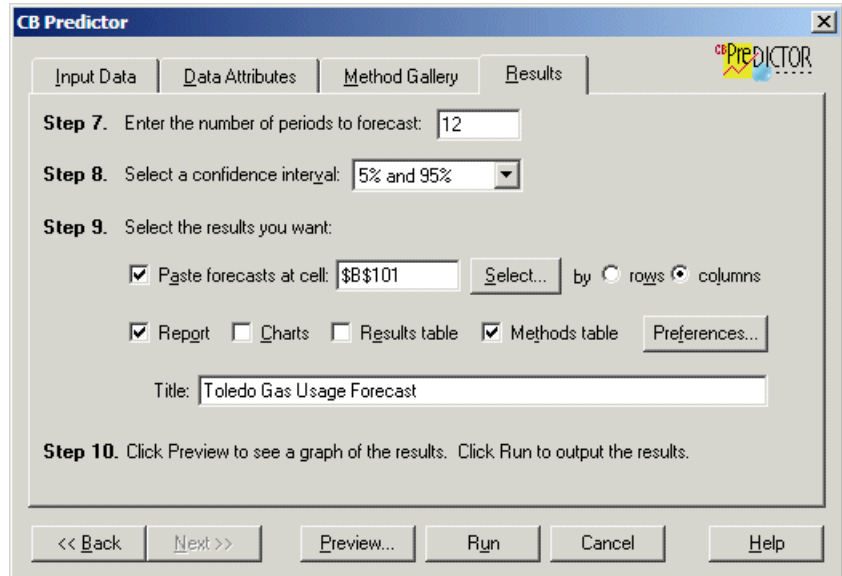
10. Click Next.

The Method Gallery tab appears.

11. Click Next again.

The Results tab appears.

Selecting results



The screenshot shows the 'CB Predictor' dialog box with the 'Results' tab selected. The dialog has four tabs: 'Input Data', 'Data Attributes', 'Method Gallery', and 'Results'. The 'Results' tab contains the following steps and options:

- Step 7.** Enter the number of periods to forecast: 12
- Step 8.** Select a confidence interval: 5% and 95%
- Step 9.** Select the results you want:
 - ☒ Paste forecasts at cell: \$B\$101 [Select...](#) by ☐ rows ☒ columns
 - ☒ Report ☐ Charts ☐ Results table ☒ Methods table [Preferences...](#)
 - Title: Toledo Gas Usage Forecast
- Step 10.** Click Preview to see a graph of the results. Click Run to output the results.

At the bottom of the dialog are five buttons: '<< Back', 'Next >>', 'Preview...', 'Run', and 'Cancel'. A 'Help' button is also present in the bottom right corner.

Figure 1.13 Results tab, set as described below

CB Predictor has five ways you can output your results:

Paste Forecast	Pastes the forecasted values to the end of your historical data.
Charts	Graphs historical, fitted, and forecasted data values, and a confidence interval.
Report	Organizes and displays summary information, forecast values and a confidence interval, charts, and method information for any or all of your data series.
Results Table	Creates a table with all the forecasted values, fitted data, and a confidence interval.
Method Table	Creates a table listing all the methods tried, the error values and statistics for each, and the parameters for each.

Paste Forecast and Charts are the most common results. The other settings provide more detailed output that includes information you might use for presentations or to investigate various forecasting methods.

This tutorial produces more detailed output. To continue with the tutorial:

12. Under Step 7, forecast the monthly usage for the next year by entering 12 in the field.

13. Select the Paste, Report, and Methods Table result settings.

14. In the Title field, type “Toledo Gas Usage Forecast.”

Now, the dialog looks like Figure 1.13 on page 21.

15. Click Preferences.

The Preferences dialog appears as shown in Figure 1.14.

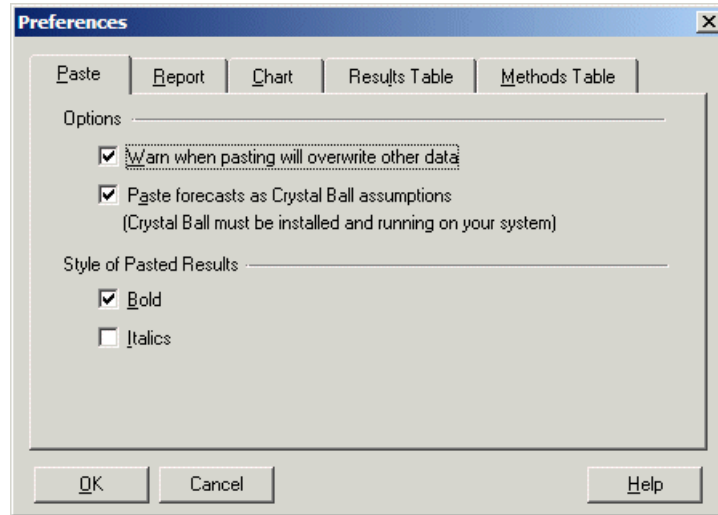


Figure 1.14 Preferences dialog

Notice that Paste Forecasts... is checked. This prepares forecasted independent variable data for use as assumptions in Crystal Ball risk analysis simulations.

16. Select the Report tab.

The Report preferences appear.

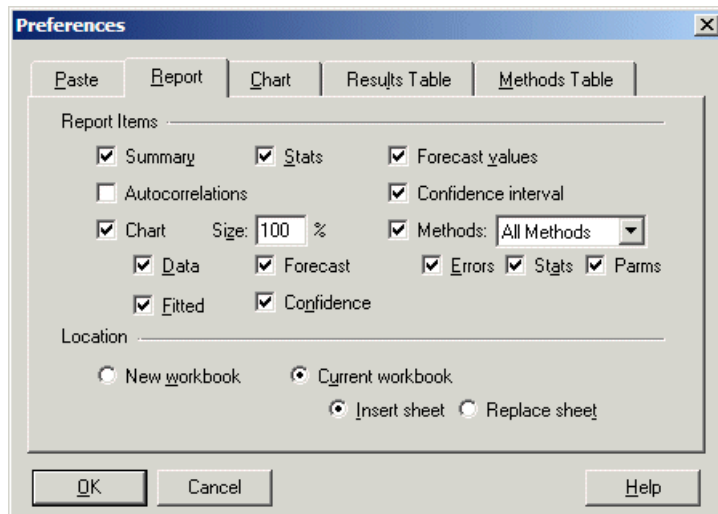


Figure 1.15 Report Preferences dialog

17. Under Methods, select Three Best from the list and unselect the Parms (Parameters) setting.

This removes the method parameters from the report.

18. Click OK.

The Results tab reappears.

19. Click Preview.

The Preview Forecast dialog appears.

Previewing results

Before you output results to your spreadsheet, you can preview the forecasted values and the best method selected with the Preview Forecast dialog.

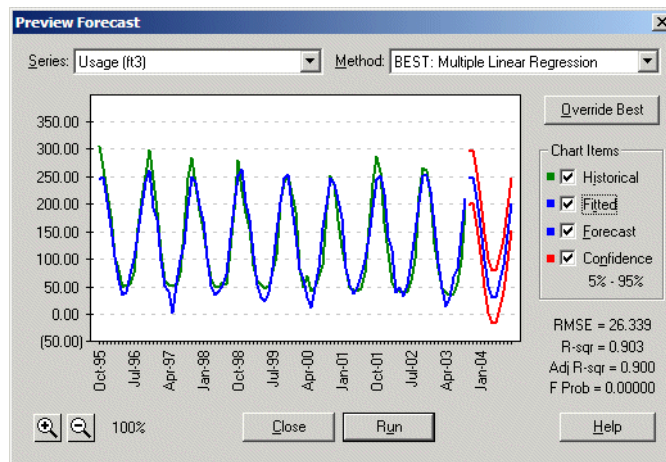


Figure 1.16 Preview Forecast dialog

In the Preview Forecast dialog, you can preview the forecasted values for all the data series and all the methods run for each. After viewing the forecast values for each method, you can also override the selected method to use for the final forecasted values.

To continue with your tutorial:

20. Select Average Temperature from the Series list in the upper left corner of the Preview Forecast dialog.

Forecasted values appear for Average Temperature. Seasonal Additive is identified as the best-fitting method as in Figure 1.17.

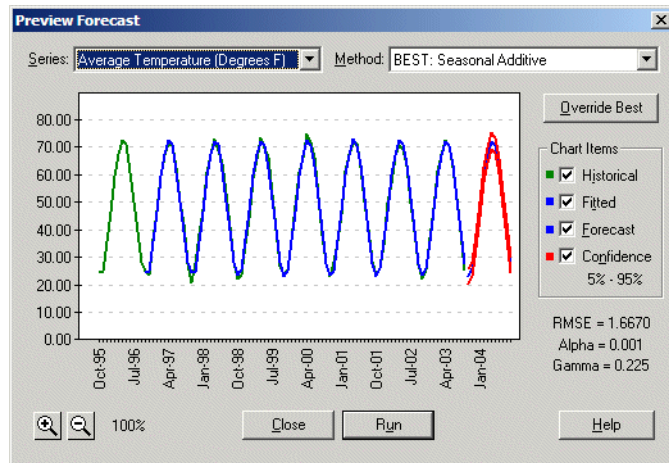


Figure 1.17 Average temperature before Method override

21. Select Single Exponential Smoothing from the Method list.

The preview changes to show the forecast using single exponential smoothing instead of the seasonal additive method.

22. Click Override Best.

This actually changes the forecast to use single exponential smoothing instead of the seasonal additive method as shown in Figure 1.18.

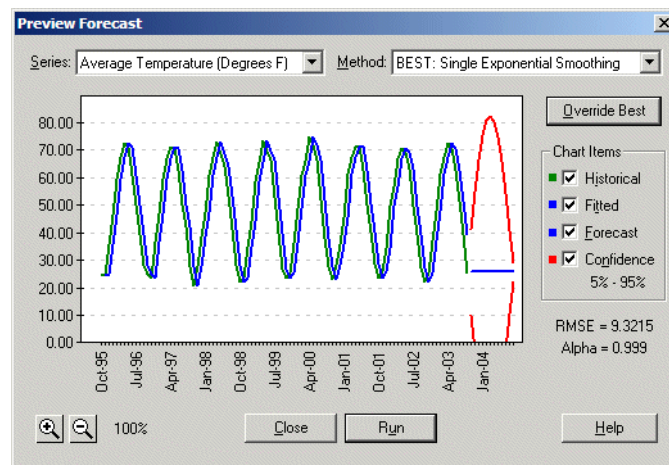
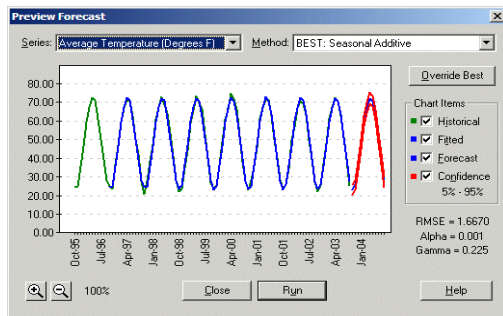


Figure 1.18 Average temperature after Method override

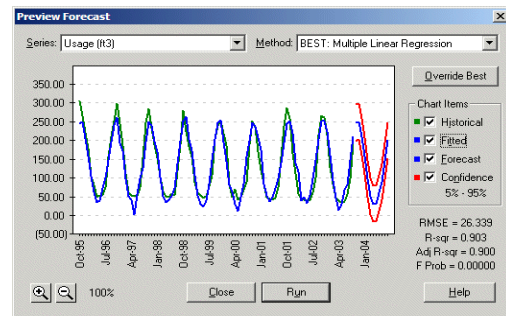
Summary

CB Predictor's primary job is to create forecasts based on your historical data. The program's method selections appear in the Preview dialog. When you override the program's selections, you should carefully analyze your results.

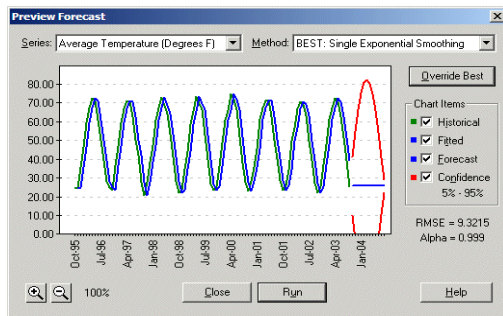
In this example, there were three independent variables that combined to forecast the dependent variable, Usage. In the tutorial, you overrode the Average Temperature reading to use the method Single Exponential Smoothing instead of Seasonal Additive. What was the effect?



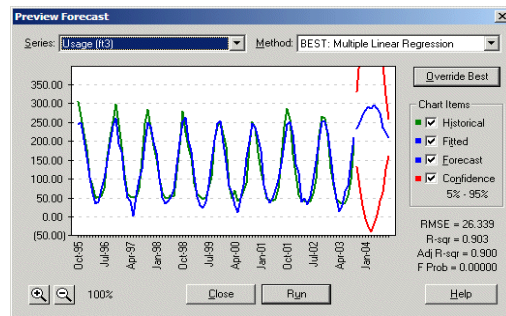
Average Temperature Before Override



Usage Before Override



Average Temperature After Override



Usage After Override

Figure 1.19 Comparison of previews before and after Average Temperature Method override

Overriding the Average Temperature had a noticeable effect on the forecast (but not the fit) of the Usage variable.

23. Click Run.

CB Predictor creates:

- The results pasted at the end of your historical data
- On a separate worksheet of your workbook, a report listing all the details of the time-series forecasts of each independent variable and the multiple linear regression of the dependent variable
- On a separate worksheet of your workbook, a Methods Table page (as *PivotTables*) listing all the time-series forecasting methods tried, their parameters, the error measures and statistics for each, and the regression parameters and statistics

Glossary Term:
PivotTable—
 An interactive table in Excel. You can move rows and columns, and filter PivotTable data.

24. Look at the results pasted below the historical data as shown in Figure 1.20.

The pane was frozen beneath the column headers so they would appear in this figure.

	Date	Usage (ft ³)	Occupancy Permits	Average Temperature (Degrees F)	Cost of Natural Gas per ccf (Dollars)
96	May-03	36.00	764	70.81	10.37
97	Jun-03	37.00	694	62.53	9.12
98	Jul-03	56.00	312	51.72	7.24
99	Aug-03	97.00	269	39.32	6.81
100	Sep-03	194.00	147	25.64	6.71
101	Oct-03	232.57	354	25.66	5.83
102	Nov-03	250.47	495	25.66	6.65
103	Dec-03	268.54	652	25.66	6.35
104	Jan-04	282.61	763	25.66	6.97
105	Feb-04	292.62	856	25.66	6.24
106	Mar-04	286.37	779	25.66	8.25
107	Apr-04	296.66	863	25.66	8.46
108	May-04	287.91	780	25.66	9.22
109	Jun-04	274.47	673	25.66	8.70
110	Jul-04	234.46	349	25.66	7.54
111	Aug-04	224.11	275	25.66	6.46
112	Sep-04	209.61	145	25.66	7.09
113					
114	Coefficient	0.00	0	-5.69	1.46
115					

Figure 1.20 Gas service predictions for the next twelve months

Notice that the independent variables have been defined as Crystal Ball assumptions with normal distributions.

25. Activate the Report worksheet and scroll to the section on the Average Temperature variable as shown in Figure 1.21.

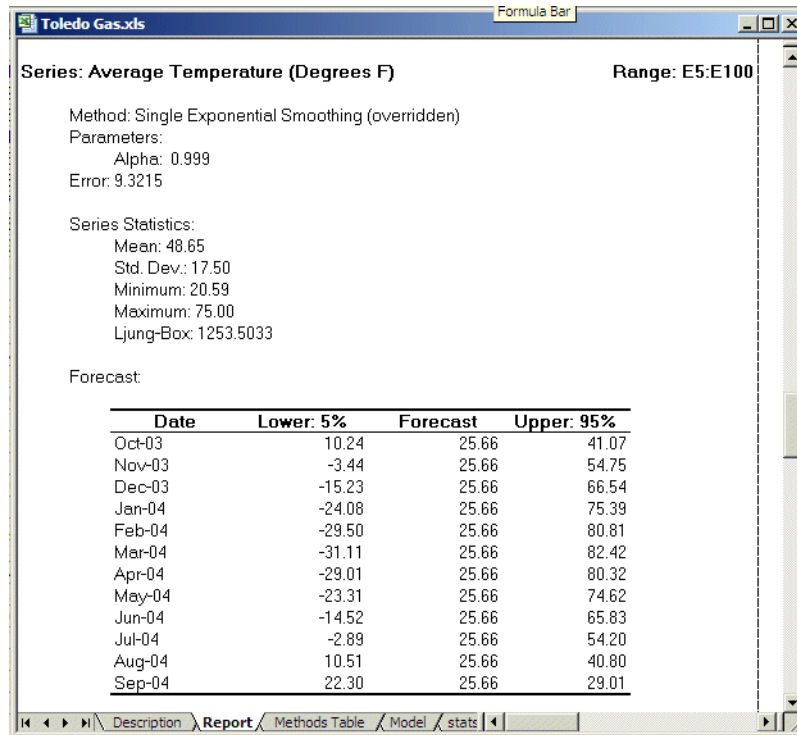


Figure 1.21 Average temperature data for Toledo Gas

Notice the indication that the method used was an override of the best method.

26. Activate the Methods Table worksheet.

Methods Table for Toledo Gas Usage Forecast
 Created: 7/22/2004 at 3:24:46 PM
 Note: dependent variables are not included
 Note: variables not forecasted are not included

Series	Occupancy Permits					
Table Items						
Methods	Rank	RMSE	MAD	MAPE	Durbin-Watson	T
Double Exponential Smoothing	7	175.97	144.64	32.636		2.131
Double Moving Average	8	176.66	149.33	35.33		1.595
Holt-Winters' Additive	2	75.336	49.407	8.628		1.707
Holt-Winters' Multiplicative	1	75.268	49.547	8.644		1.711
Seasonal Additive	3	75.338	49.405	8.644		1.706
Seasonal Multiplicative	4	75.338	49.419	8.655		1.706
Single Exponential Smoothing	6	167.02	140.51	33.407		1.527
Single Moving Average	5	166.98	140.44	33.388		1.529

Figure 1.22 Toledo Gas methods table, default view

27. Next to the Series button, select Average Temperature.

The table changes to show the parameters and statistics for each method of the Average Temperature forecast as shown in Figure 1.23.

Methods Table for Toledo Gas Usage Forecast
 Created: 7/20/2004 at 3:43:18 PM
 Note: dependent variables are not included
 Note: variables not forecasted are not included

Series	Average Temperature (Degrees F)					
Table Items						
Methods	Rank	RMSE	MAD	MAPE		
Double Exponential Smoothing	5	6.0181	4.9758	12		
Double Moving Average	6	8.278	7.0697	16		
Holt-Winters' Additive	2	1.667	1.3188	3		
Holt-Winters' Multiplicative	4	1.6672	1.3188	3		
Seasonal Additive	1	1.667	1.3188	3		
Seasonal Multiplicative	3	1.6672	1.3188	3		
Single Exponential Smoothing	8	9.3215	8.3387	19		
Single Moving Average	7	9.3142	8.332	19		

Figure 1.23 Methods table with average temperature selected

28. Click the Series button and drag it to the left of the Methods button.

The method table expands to include all the data series. When you drop the Series button next to the Methods button, the list of methods repeats for each series.

Series	Methods	Rank	RMSE	MAD	MAPE
Average Temperature (Degrees F)	Double Exponential Smoothing	5	6.0181	4.9758	12.1
	Double Moving Average	6	8.278	7.0697	16.9
	Holt-Winters' Additive	2	1.667	1.3188	3.0
	Holt-Winters' Multiplicative	4	1.6672	1.3188	3.0
	Seasonal Additive	1	1.667	1.3188	3.0
	Seasonal Multiplicative	3	1.6672	1.3188	3.0
	Single Exponential Smoothing	8	9.3215	8.3387	19.6
Cost of Natural Gas per ccf (Dollars)	Single Moving Average	7	9.3142	8.332	19.6
	Double Exponential Smoothing	8	1.4395	1.146	15.9
	Double Moving Average	6	1.2967	1.049	14.5
	Holt-Winters' Additive	2	1.2583	0.9595	13.7
	Holt-Winters' Multiplicative	4	1.2586	0.9595	13.8
	Seasonal Additive	1	1.2583	0.9595	13.7
	Seasonal Multiplicative	3	1.2586	0.9595	13.8

Figure 1.24 Methods grouped by series

29. Click the down arrow beside the Table Items button.

A list of fields appears.

30. Uncheck all the items except for Rank.**31. Click OK.**

The methods table changes to show one parameter: Rank. Look at the Average Temperature data. Under Rank, Single Exponential Smoothing is highlighted in blue, bold text to show that it was used to generate the results. Seasonal Additive, originally the best, is still listed with a Rank equal to 1.

Series	Methods	Rank
Average Temperature (Degrees F)	Double Exponential Smoothing	5
	Double Moving Average	6
	Holt-Winters' Additive	2
	Holt-Winters' Multiplicative	4
	Seasonal Additive	1
	Seasonal Multiplicative	3
	Single Exponential Smoothing	8
Cost of Natural Gas per ccf (Dollars)	Double Exponential Smoothing	8
	Double Moving Average	6
	Holt-Winters' Additive	2
	Holt-Winters' Multiplicative	4
	Seasonal Additive	1
	Seasonal Multiplicative	3
	Occupancy Permits	7

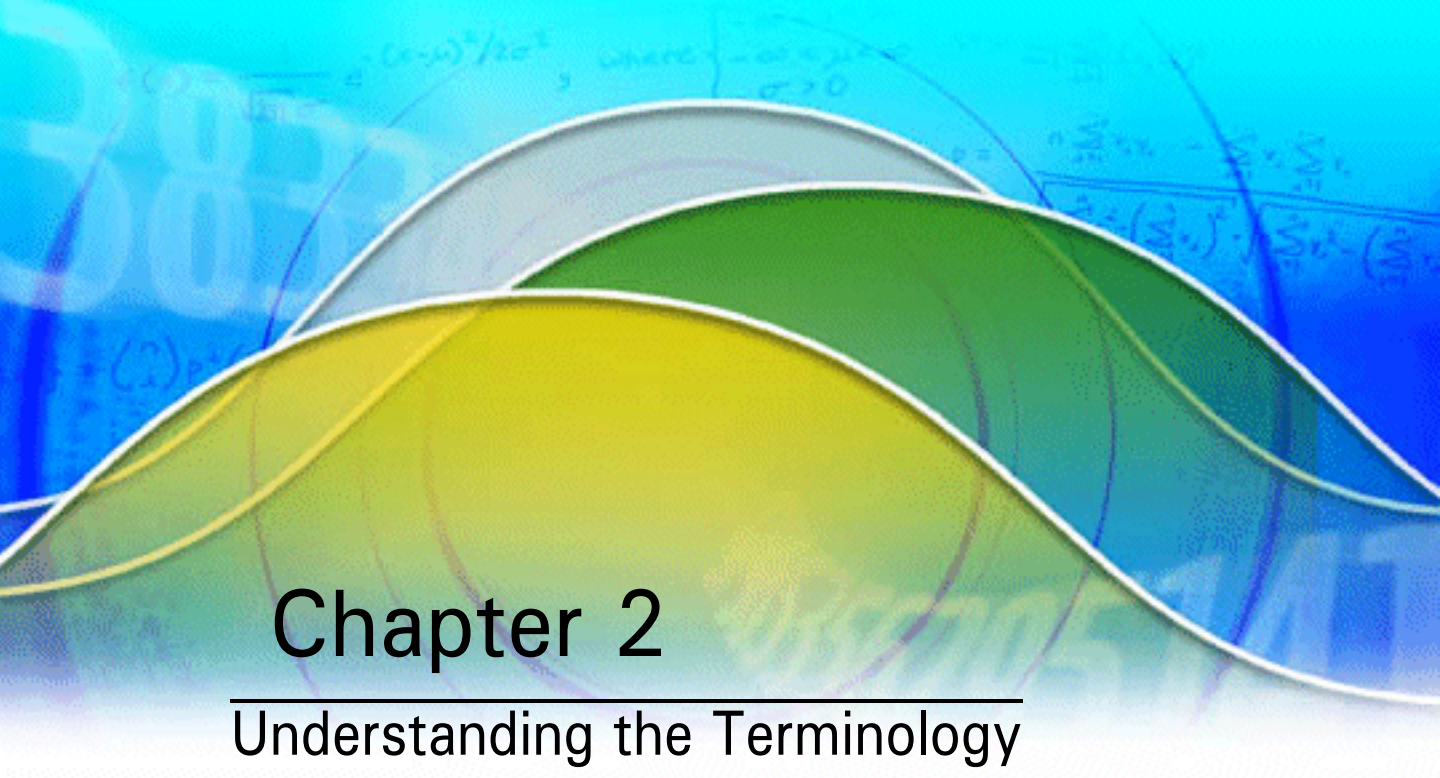
Figure 1.25 Methods within each series identified by rank

32. Move the Methods button to the left of the Series button.

The PivotTable reorganizes to show all the series grouped by method type as shown in Figure 1.26.

Methods	Series	Rank
Double Exponential Smoothing	Average Temperature (Degrees F)	5
	Cost of Natural Gas per ccf (Dollars)	8
	Occupancy Permits	7
Double Moving Average	Average Temperature (Degrees F)	6
	Cost of Natural Gas per ccf (Dollars)	6
	Occupancy Permits	8
Holt-Winters' Additive	Average Temperature (Degrees F)	2
	Cost of Natural Gas per ccf (Dollars)	2
	Occupancy Permits	2
Holt-Winters' Multiplicative	Average Temperature (Degrees F)	4
	Cost of Natural Gas per ccf (Dollars)	4
	Occupancy Permits	1
Seasonal Additive	Average Temperature (Degrees F)	1
	Cost of Natural Gas per ccf (Dollars)	1

Figure 1.26 Series grouped within methods



Chapter 2

Understanding the Terminology

In this chapter

- Forecasting
- Time-series forecasting
- Time-series forecasting error measures
- Time-series forecasting techniques
- Time-series forecasting statistics
- Multiple linear regression
- Regression methods
- Regression statistics
- Historical data statistics

This chapter describes forecasting terminology. It defines the time-series forecasting methods that CB Predictor uses, as well as other forecasting-related terminology.

This chapter also describes the statistics the program generates and the techniques that CB Predictor uses to do the calculations and select the best-fitting method.

Forecasting

Forecasting refers to the act of predicting the future, usually for purposes of planning and managing resources. There are many scientific approaches to forecasting. You can perform “what-if” forecasting by creating a model and simulating outcomes, as with Crystal Ball, or by collecting data over a period of time and analyzing the trends and patterns. CB Predictor uses the latter concept, analyzing the patterns of a time series to forecast future data.

The scientific approaches to forecasting usually fall into one of several categories:

Time-series	Performs time-series analysis on past patterns of data to forecast results. This works best for stable situations where conditions are expected to remain the same.
Regression	Forecasts results using past relationships between a variable of interest and several other variables that might influence it. This works best for situations where you need to identify the different effects of different variables. This category includes multiple linear regression.
Simulation	Randomly generates many different scenarios for a model to forecast the possible outcomes. This method works best where you might not have historical data, but you can build the model of your situation to analyze its behavior.
Qualitative	Uses subjective judgment and expert opinion to forecast results. These methods work best for situations for which there are no historical data or models available.

CB Predictor uses both time-series and multiple linear regression for forecasting. Crystal Ball uses simulation. Each technique and method has advantages and disadvantages for particular types of data, so often you might forecast your data using several methods and then select the method that yields the best results.

Time-series forecasting

Glossary Term:
level—

A starting point for the forecast.

Glossary Term:
trend—

A long-term increase or decrease in time-series data.

Glossary Term:
seasonality—

The change that seasonal factors cause in a data series. For example, if sales increase during the Christmas season but not during the summer, the data is seasonal with a one-year period.

Time-series forecasting assumes historical data is a combination of a pattern and some random error. Its goal is to isolate the pattern from the error by understanding the pattern's *level*, *trend*, and *seasonality*. You can then measure the error using a statistical measurement to describe both how well a pattern reproduces historical data and to estimate how accurately it forecasts the data into the future. For more information on these error measurements, see “Time-series forecasting error measures” on page 42.

When you select different forecasting methods from the Methods Gallery, CB Predictor tries all of them. It then ranks them according to which method has the lowest error, depending on the error measure selected in the Advanced dialog. The method with the lowest error is the best method.

There are two primary techniques of time-series forecasting used in CB Predictor. They are:

Nonseasonal smoothing	Estimates a trend by removing extreme data and reducing data randomness.
Seasonal smoothing	Combines smoothing data with an adjustment for seasonal behavior.

Nonseasonal smoothing methods

Smoothing models attempt to forecast by removing extreme changes in past data. The following methods are available.

Single moving average

Smooths out historical data by averaging the last several periods and projecting the last average value forward. CB Predictor can automatically calculate the optimal number of periods to average, or you can select the number of periods to average.

This method is best for volatile data with no trend or seasonality. It results in a straight, flat-line forecast.

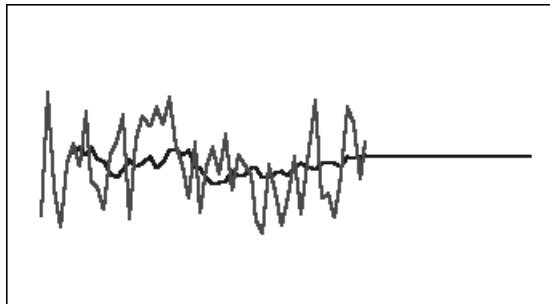


Figure 2.1 Typical single moving average data, fit, and forecast line

Double moving average

Applies the moving average technique twice, once to the original data and then to the resulting single moving average data. This method then uses both sets of smoothed data to project forward. CB Predictor can automatically calculate the optimal number of periods to average, or you can select the number of periods to average.

This method is best for historical data with a trend but no seasonality. It results in a straight, sloped-line forecast.

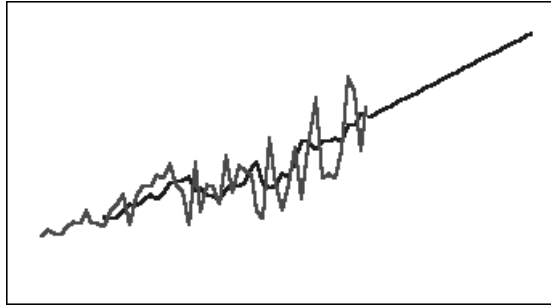


Figure 2.2 Typical double moving average data, fit, and forecast line

CB Predictor Note: For See Appendix B, “Time-Series Forecasting Method Formulas” for more information about the formulas CB Predictor uses for the following methods.

Single exponential smoothing (SES)

Weights all of the past data with exponentially decreasing weights going into the past. In other words, usually the more recent data have greater weight. This largely overcomes the limitations of moving averages or percentage change models. CB Predictor can automatically calculate the optimal smoothing constant, or you can manually define the smoothing constant.

This method is best for volatile data with no trend or seasonality. It results in a straight, flat-line forecast.

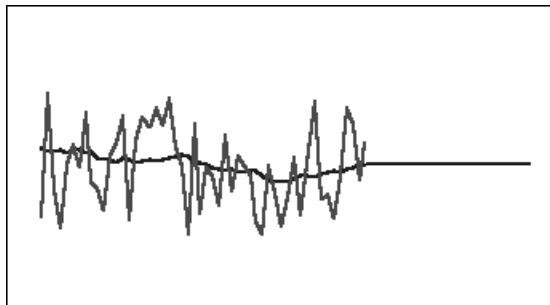


Figure 2.3 Typical single exponential smoothing data, fit, and forecast line

Holt’s double exponential smoothing (DES)

Double exponential smoothing applies SES twice, once to the original data and then to the resulting SES data. CB Predictor uses Holt’s method for

double exponential smoothing, which can use a different parameter for the second application of the SES equation. CB Predictor can automatically calculate the optimal smoothing constants, or you can manually define the smoothing constants.

This method is best for data with a trend but no seasonality. It results in a straight, sloped-line forecast.

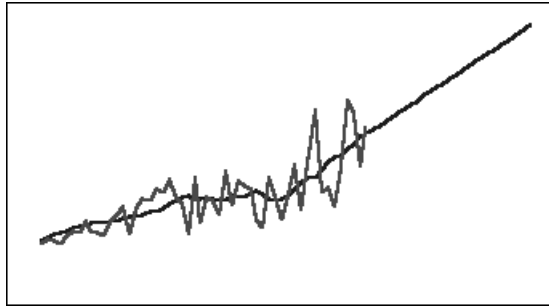


Figure 2.4 Typical double exponential smoothing data, fit, and forecast line

Nonseasonal smoothing parameters

There are several smoothing parameters used by the nonseasonal methods. For the moving average methods, the formulas use one parameter: period. When performing a moving average, you average over a number of periods. For single moving average, the number of periods can be any whole number between 1 and half the number of data points. For double moving average, the number of periods can be any whole number between 2 and one-third the number of data points.

For single exponential smoothing, there is one parameter: alpha. Alpha (α) is the smoothing constant. The value of alpha can be any number between 0 and 1, not inclusive.

For Holt's double exponential smoothing, there are two parameters: alpha and beta. Alpha is the same smoothing constant as described above for single exponential smoothing. Beta (β) is also a smoothing constant exactly like alpha except that it is used during second smoothing. The value of beta can be any number between 0 and 1, not inclusive.

Seasonal smoothing methods

Seasonal exponential smoothing methods extend the simple exponential smoothing methods by adding an additional component to capture the seasonal behavior of the data. There are four seasonal exponential smoothing methods used in CB Predictor.

CB Predictor Note: See Appendix B, “Time-Series Forecasting Method Formulas” for more information about the formulas CB Predictor uses.

Seasonal Additive Smoothing

Calculates a seasonal index for historical data that don’t have a trend. The method produces exponentially smoothed values for the level of the forecast and the seasonal adjustment to the forecast. The seasonal adjustment is added to the forecasted level, producing the Seasonal Additive forecast.

This method is best for data without trend but with seasonality that doesn’t increase over time. It results in a curved forecast that reproduces the seasonal changes in the data.

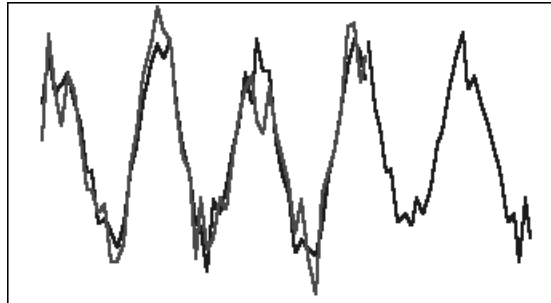


Figure 2.5 Typical data, fit, and forecast curve for seasonal additive smoothing without trend

Seasonal Multiplicative Smoothing

Calculates a seasonal index for historical data that don’t have a trend. The method produces exponentially smoothed values for the level of the forecast and the seasonal adjustment to the forecast. The seasonal adjustment is multiplied by the forecasted level, producing the Seasonal Multiplicative forecast.

This method is best for data without trend but with seasonality that increases or decreases over time. It results in a curved forecast that reproduces the seasonal changes in the data.

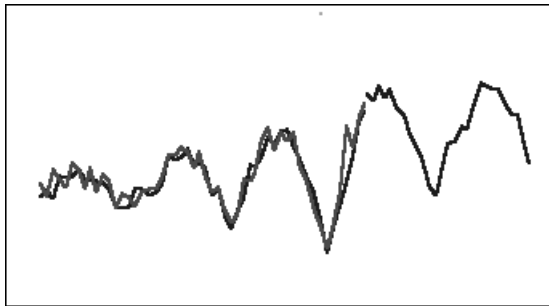


Figure 2.6 Typical data, fit, and forecast curve for seasonal multiplicative smoothing without trend

Holt-Winters' Additive Seasonal Smoothing

Is an extension of Holt's exponential smoothing that captures seasonality. This method is based upon three equations that can be found in Appendix B. The method produces exponentially smoothed values for the level of the forecast, the trend of the forecast, and the seasonal adjustment to the forecast. This seasonal additive method adds the seasonality factor to the trended forecast, producing the Holt-Winters' Additive forecast.

This method is best for data with trend and seasonality that doesn't increase over time. It results in a curved forecast that shows the seasonal changes in the data.

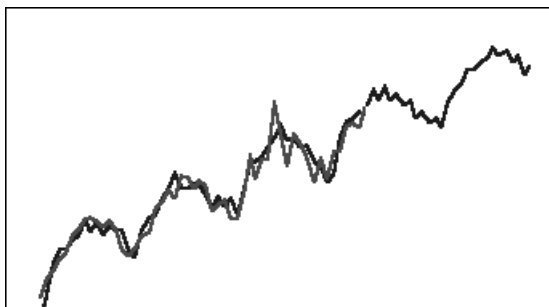


Figure 2.7 Typical data, fit, and forecast curve for Holt-Winters' additive seasonal smoothing

Holt-Winters' Multiplicative Seasonal Smoothing

Is similar to the Holt-Winters' Additive method. This method also calculates exponentially smoothed values for level, trend, and seasonal adjustment to the forecast. This method's equations can also be found in Appendix B. This seasonal multiplicative method multiplies the trended forecast by the seasonality, producing the Holt-Winters' Multiplicative forecast.

This method is best for data with trend and with seasonality that increases over time. It results in a curved forecast that reproduces the seasonal changes in the data.

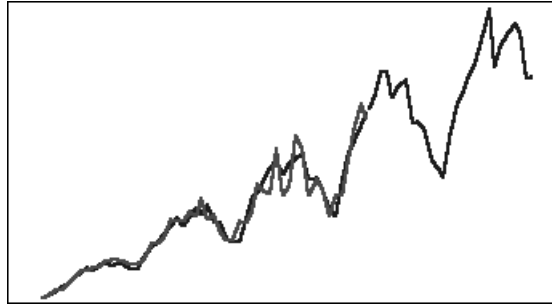


Figure 2.8 Typical data, fit, and forecast curve for Holt-Winters' multiplicative seasonal smoothing

Seasonal smoothing parameters

There are three smoothing parameters used by the seasonal methods: alpha, beta, and gamma.

alpha (α)	Smoothing parameter for the level component of the forecast. The value of alpha can be any number between 0 and 1, not inclusive.
beta (β)	Smoothing parameter for the trend component of the forecast. The value of beta can be any number between 0 and 1, not inclusive.
gamma (γ)	Smoothing parameter for the seasonality component of the forecast. The value of gamma can be any number between 0 and 1, not inclusive.

Each seasonal method uses some or all of these parameters, depending on the forecasting method. For example, the seasonal additive smoothing method doesn't account for trend, so it doesn't use the beta parameter.

Time-series forecasting error measures

One component of every time-series forecast is the data's random error that is not explained by the forecast formula or by the trend and seasonal patterns. The error is measured by fitting points for the time periods with historical data, and then comparing the fitted points to the historical data.

All the examples are based on the set of data illustrated in the chart below. Most of the formulas refer to the actual points (Y) and the fitted points (\hat{Y}_t). In the chart below, the horizontal axis illustrates the time periods (t) and the vertical axis illustrates the data point values.

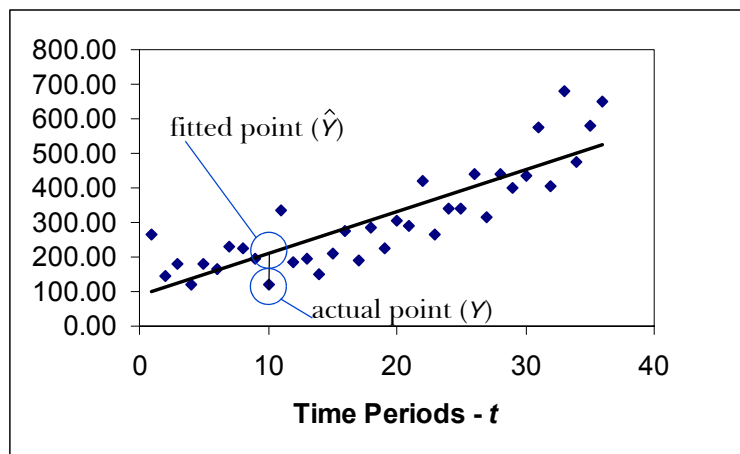


Figure 2.9 Sample data

CB Predictor measures the error using one of the following methods:

- RMSE, below
- MAD, page 43
- MAPE, page 43

RMSE

RMSE (root mean squared error) is an absolute error measure that squares the deviations to keep the positive and negative deviations from cancelling out each other. This measure also tends to exaggerate large errors, which can help eliminate methods with large errors.

MAD

MAD (mean absolute deviation) is an absolute error measure that originally became very popular (in the days before hand-held calculators) because it didn't require the calculation of squares or square roots. While it is still fairly reliable and widely used, it is most accurate for normally distributed data.

MAPE

MAPE (mean absolute percentage error) is a relative error measure that uses absolute values. There are two advantages of this measure. First, the absolute values keep the positive and negative errors from cancelling out each other. Second, because relative errors don't depend on the scale of the dependent variable, this measure lets you compare forecast accuracy between differently scaled time-series data.

Time-series forecasting techniques

CB Predictor uses one of four forecasting techniques to perform time-series forecasting: standard, simple lead, weighted lead, and holdout.

Standard forecasting

Standard forecasting optimizes the forecasting parameters to minimize the error measure between the fit values and the historical data for the same period. For example, if your historical data were:

Period	1	2	3	4	5	6	7
Value	472	599	714	892	874	896	890

And your fit data were:

Period	1	2	3	4	5	6	7
Value	488	609	702	888	890	909	870

CB Predictor calculates the RMSE using the differences between the historical data and the fit data from the same periods. For example:

$$(472-488)^2 + (599-609)^2 + (714-702)^2 + (892-888)^2 + \dots$$

For standard forecasting, CB Predictor optimizes the forecasting parameters so that the RMSE calculated in this way is minimized.

Simple lead forecasting

Simple lead forecasting optimizes the forecasting parameters to minimize the error measure between the historical data and the fit values, offset by a specified number of periods (lead). Use this forecasting technique when a forecast for some future time period has the greatest importance, more so than the forecasts for the previous or later periods.

For example, your company must order extremely expensive manufacturing components two months in advance, making any forecast for two months out the most important. For the same historical and fit example data described above in Standard Forecasting, CB Predictor calculates the RMSE using the difference between the historical data and the fit data from an offset number of periods (lead). With a lead of 2, the differences used in the RMSE calculation are:

$$(472-702)^2 + (599-888)^2 + (714-890)^2 + (892-909)^2 + \dots$$

For simple lead forecasting, CB Predictor optimizes the forecasting parameters so that the RMSE calculated in this way is minimized.

Weighted lead forecasting

Weighted lead forecasting optimizes the forecasting parameters to minimize the *average* error measure between the historical data and the fit values, offset by 0, 1, 2, etc., up to the specified number of periods (weighted lead). It uses the simple lead technique for several lead periods and then averages the forecast over the periods, optimizing this average value. Use this technique when the future forecast for several periods is most important.

For example, your company must order extremely expensive manufacturing components zero, one, and two months in advance, making any forecast for all the time periods up to two months out the most important. For the same historical and fit example data described above in Standard Forecasting, CB Predictor calculates the RMSE for each lead up to the weighted lead using the difference between the historical data and the fit data from a set of offset periods (individual leads). With a weighted lead of 2, the differences used in the RMSE calculations are:

$$(472-488)^2 + (599-609)^2 + (714-702)^2 + \dots \text{ (lead of 0)}$$

$$(472-609)^2 + (599-702)^2 + (714-888)^2 + \dots \text{ (lead of 1)}$$

$$(472-702)^2 + (599-888)^2 + (714-890)^2 + \dots \text{ (lead of 2)}$$

Then CB Predictor averages the RMSE for the lead of 0, the lead of 1, and the lead of 2. For weighted lead forecasting, CB Predictor optimizes the forecasting parameters to minimize the average of the RMSE calculations.

Holdout

Holdout forecasting:

1. **Removes the last few data points of your historical data.**
2. **Calculates the fit and forecast points using the remaining historical data.**
3. **Compares the error between the forecasted points and their corresponding, excluded, historical data points.**
4. **Changes the parameters to minimize the error between the forecasted points and the excluded points.**

CB Predictor determines the optimal forecast parameters using only the non-holdout set of data.

CB Predictor Note: *If you have a small amount of data and want to use seasonal forecasting methods, using the holdout technique might restrict you to nonseasonal methods.*

Statistical Note: *For more information on the holdout technique and when to use it effectively, see the Makridakis, Wheelwright, and Hyndman reference in the bibliography.*

Time-series forecasting statistics

Theil's *U*

Glossary Term:
naïve forecast—
A forecast obtained
with minimal effort
based only on the most
recent data; e.g., using
the last data point to
forecast the next
period.

Theil's *U* statistic is a relative error measure that compares the forecasted results with a *naïve forecast*. It also squares the deviations to give more weight to large errors and to exaggerate errors, which can help eliminate methods with large errors. For the formula, see page 137.

Table 2.1 Interpreting Theil's *U*

<i>Theil's U statistic</i>	<i>Means:</i>
Less than 1	The forecasting technique is better than guessing.
1	The forecasting technique is about as good as guessing.
More than 1	The forecasting technique is worse than guessing.

Multiple linear regression

Multiple linear regression is used for data where one data series (the dependent variable) is a function of, or depends on, other data series (the independent variables). For example, the yield of a lettuce crop depends on the amount of water provided, the hours of sunlight each day, and the amount of fertilizer used.

The goal of multiple linear regression is to find an equation that most closely matches the historical data. The word “multiple” indicates that you can use more than one independent variable to define your dependent variable in the regression equation. The word “linear” indicates that the regression equation is a *linear equation*.

Glossary Term:
linear equation—
An equation with only
linear terms. A linear
equation has no terms
containing variables
with exponents or
variables multiplied by
each other

The linear equation describes how the independent variables (*x*1, *x*2, *x*3,...) combine to define the single dependent variable (*y*). Multiple linear regression finds the coefficients for the equation:

$$y = b0 + b1x1 + b2x2 + b3x3 + ... + e$$

where *b*1, *b*2, and *b*3, are the coefficients of the independent variables, *b*0 is the *y*-intercept, and *e* is the error.

If there is only one independent variable, the equation defines a straight line. This uses a special case of multiple linear regression called simple linear regression, with the equation:

$$y = b_0 + b_1x + e$$

where b_0 is where on the graph the line crosses the y axis, x is the independent variable, and e is the error. When the regression equation has only two independent variables, it defines a plane. When the regression equation has more than two independent variables, it defines a *hyperplane*.

To find the coefficients of these equations, CB Predictor uses *singular value decomposition*. For more information on this technique, see “Singular value decomposition” on page 143.

Glossary Term:
hyperplane—
A geometric plane that spans more than two dimensions.

Glossary Term:
singular value decomposition—
A method that solves a set of equations for the coefficients of a regression equation.

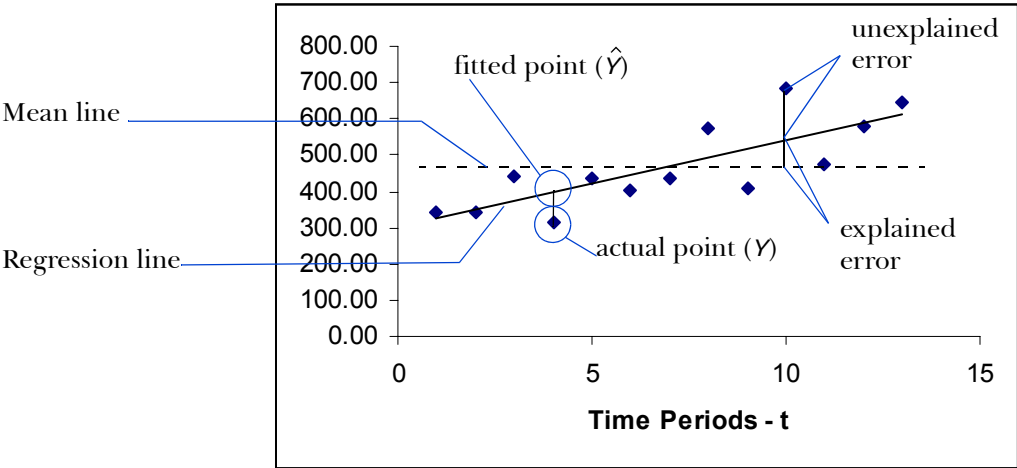


Figure 2.10 Parts of a scatter plot

Regression methods

CB Predictor uses one of three regression methods: standard regression, forward stepwise regression, and iterative stepwise regression.

Standard regression

Standard regression performs multiple linear regression, generating regression coefficients for each independent variable you specify, no matter how significant.

Forward stepwise regression

Forward stepwise regression adds one independent variable at a time to the multiple linear regression equation, starting with the independent variable with the most significant probability of the correlation (partial F statistic). It then recalculates the partial F statistic for the remaining independent variables, taking the existing regression equation into consideration.

CB Predictor Note: *The resulting multiple linear regression equation will always have at least one independent variable.*

Forward stepwise regression continues to add independent variables until either:

- It runs out of independent variables.
- It reaches one of the selected stopping criteria in the Stepwise Options dialog.
- The number of included independent variables reaches one-third the number of data points in the series.

There are two stopping criteria:

R -squared	Stops the stepwise regression if the difference between a specified statistic (either R^2 or adjusted R^2) for the previous and new regression solutions is below a threshold value. When this happens, CB Predictor does not use the last independent variable.
--------------	---

For example, the third step of a stepwise regression results in an R^2 value of 0.81, and the fourth step adds another independent variable and results in an R^2 value is 0.83. The difference between the R^2 values is 0.02.

If the Threshold value is 0.03, CB Predictor returns to the regression equation for the third step and stops the stepwise regression.

F-test
significance

Stops the stepwise regression if the probability of the *F* statistic for a new solution is above a maximum value.

For example, if you set the maximum probability to 0.05 and the *F* statistic for the fourth step of a stepwise regression results in a probability of 0.08, CB Predictor returns to the regression equation for the third step and stops the stepwise regression.

Iterative stepwise regression

Iterative stepwise regression adds or removes one independent variable at a time to or from the multiple linear regression equation.

To perform iterative stepwise regression, CB Predictor:

- 1. Calculates the partial *F* statistic for each independent variable.**
- 2. Adds the independent variable with the most significant correlation (partial *F* statistic).**
- 3. Checks the partial *F* statistic of the independent variables in the regression equation to see if any became insignificant (have a probability below the minimum) with the addition of the latest independent variable.**
- 4. Removes the least significant of any insignificant independent variables one at a time.**
- 5. Repeat step 3 until no insignificant variables remain in the regression equation.**
- 6. Repeat steps 1 through 5 until:**
 - The model runs out of independent variables.
 - The regression reaches one of the stopping criteria (see the previous section for information on how the stopping criteria work).
 - The same independent variable is added and then removed.

CB Predictor Note: *The resulting equation will always have at least one independent variable.*

Regression statistics

Once CB Predictor finds the regression equation, it calculates several statistics to help you evaluate the regression. See Appendix E, “Regression Statistic Formulas” for more information on the formulas CB Predictor uses to calculate these statistics.

R^2

Coefficient of determination. This statistic indicates the percentage of the variability of the dependent variable that the regression equation explains.

For example, an R^2 of 0.36 indicates that the regression equation accounts for 36% of the variability of the dependent variable.

Adjusted R^2

Corrects R^2 to account for the degrees of freedom in the data. In other words, the more data points you have, the more universal the regression equation is. However, if you have only the same number of data points as variables, the R^2 might appear deceptively high. This statistic corrects for that.

For example, the R^2 for one equation might be very high, indicating that the equation accounted for almost all the error in the data. However, this value might be inflated if the number of data points was insufficient to calculate a universal regression equation.

SSE

Sum of square deviations. The least squares technique for estimating regression coefficients minimizes this statistic, which measures the error not eliminated by the regression line.

For any line drawn through a scatter plot of data, there are a number of different ways to determine which line fits the data best. One method is to compare the fit of lines is to calculate the SSE (sum of the squared errors) for each line. The lower the SSE, the better the fit of the line to the data.

***F* statistic**

Tests the significance of the regression equation as measured by R^2 . If this value is significant, it means that the regression equation does account for some of the variability of the dependent variable.

***t* statistic**

Tests the significance of the relationship between the coefficients of the dependent variable and the individual independent variable, in the presence of the other independent variables. If this value is significant, it means that the independent variable does contribute to the dependent variable.

p

Indicates the probability of your calculated *F* or *t* statistic being as large as it is (or larger) by chance. A low *p* value is good and means that the *F* statistic is not coincidental and, therefore, is significant. A significant *F* statistic means that the relationship between the dependent variable and the combination of independent variables is significant.

Generally, you want your *p* to be less than 0.05.

Durbin-Watson

Glossary Term:
autocorrelation—
Describes a relationship
or correlation between
values of the same data
series at different time
periods.

Glossary Term:
lag—
Defines the offset when
comparing a data series
with itself. For
autocorrelation, this
refers to the offset of
data that you choose
when correlating a data
series with itself.

Detects *autocorrelation* at *lag* 1. This means that each time-series value influences the next value. This is the most common type of autocorrelation. For the formula, see page 138.

The value of this statistic can be any value between 0 and 4. Values indicate slow-moving, none, or fast-moving autocorrelation, as shown in Table 2.2.

Table 2.2 Interpreting the Durbin Watson statistic

<i>Durbin-Watson statistic</i>	<i>Means:</i>
Less than 1	The errors are positively correlated. An increase in one period follows an increase in the previous period.
2	No autocorrelation.
More than 3	The errors are negatively correlated. An increase in one period follows an decrease in the previous period.

Avoid using independent variables that have errors with a strong positive or negative correlation, since this can lead to an incorrect forecast for the dependent variable.

Historical data statistics

CB Predictor automatically calculates the following statistics for historical data series:

mean

The mean of a set of values is found by adding the values and dividing their sum by the number of values. The term “average” usually refers to the mean. For example, 5.2 is the mean or average of 1, 3, 6, 7, and 9.

standard deviation

The standard deviation is the square root of the variance for a distribution. Like the variance, it is a measure of dispersion about the mean and is useful for describing the “average” deviation.

For example, you can calculate the standard deviation of the values 1, 3, 6, 7, and 9 by finding the square root of the variance that is calculated in the variance example below.

The standard deviation, denoted as s , is calculated from the variance as follows:

$$s = \sqrt{10.2} = 3.19$$

where the variance is a measure of the dispersion, or spread, of a set of values about the mean. When values are close to the mean, the variance is small. When values are widely scattered about the mean, the variance is larger.

To calculate the variance of a set of values:

- 1. Find the mean or average.**
- 2. For each value, calculate the difference between the value and the mean.**
- 3. Square these differences.**
- 4. Divide by n-1, where n is the number of differences.**

For example, suppose your values are 1, 3, 6, 7, and 9. The mean is 5.2. The variance, denoted by s^2 , is calculated as follows:

$$\begin{aligned}s^2 &= \frac{(1-5.2)^2 + (3-5.2)^2 + (6-5.2)^2 + (7-5.2)^2 + (9-5.2)^2}{5-1} \\ &= \frac{40.8}{4} = 10.2\end{aligned}$$

minimum

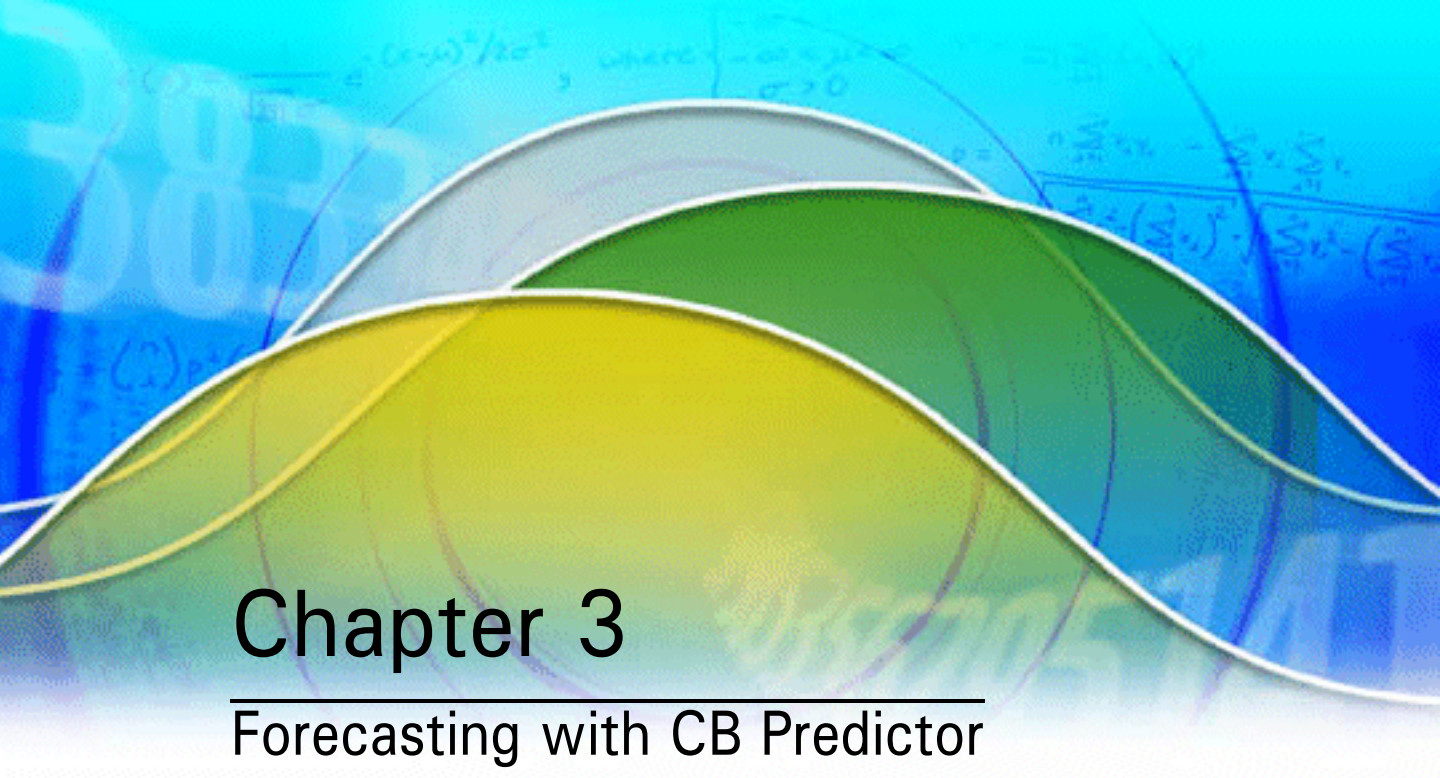
The minimum is the smallest value in the data range.

maximum

The maximum is the largest value in the data range.

Ljung-Box statistic

Measures whether a set of autocorrelations is significantly different from a set of autocorrelations that are all zero. See page 139 for the formula.



Chapter 3

Forecasting with CB Predictor

In this chapter

- Overview
- Creating a spreadsheet with historical data
- Loading and starting CB Predictor
- Guidelines for using CB Predictor
- Analyzing the results
- Customizing reports and charts

This chapter contains detailed procedures for using CB Predictor. It describes how to forecast using both time-series forecasting methods and multiple linear regression. It also describes all the settings you can choose for your results.

Overview

When you use CB Predictor for the first time, there are several things you need to do:

1. **Create an Excel spreadsheet with your historical data.**
2. **Start CB Predictor.**
3. **Run CB Predictor.**
4. **Analyze your results.**
5. **Customize your results.**

This chapter describes how to complete each of these steps to make forecasting your historical data an easy task.

Creating a spreadsheet with historical data

Before using CB Predictor, you must create an Excel spreadsheet with your historical data. Creating a spreadsheet for use with CB Predictor is easy. The spreadsheet must include:

- Optionally, a descriptive spreadsheet title.
- Optionally, a date (or other time period, such as Q2-2004) column or row, either at the top or along the left side of your data. If you format your dates as Excel dates, CB Predictor's Intelligent Input can find the dates, extend them with the forecasted values, and use them as labels on charts.
- Historical data, spaced equal time periods apart, in columns or rows adjacent to the date column or row. You can use CB Predictor to simultaneously forecast from one to 10,000 adjacent historical data series.

Excel Note: *Excel only has 256 possible columns, compared to 16,000 or 65,000 possible rows (depending on your version). So, if you have a large number of historical data points, organize your data in columns. If you have a large number of data series, organize your data in rows.*

To produce a reasonable forecast, you should have at least 6 historical data values. To use seasonal forecasting methods, you need at least two complete cycles of data.

- Optionally, headings for each data column or row, such as SKU 23442, Gas Usage, or Interest Rate.

The Toledo Gas spreadsheet has all these components.

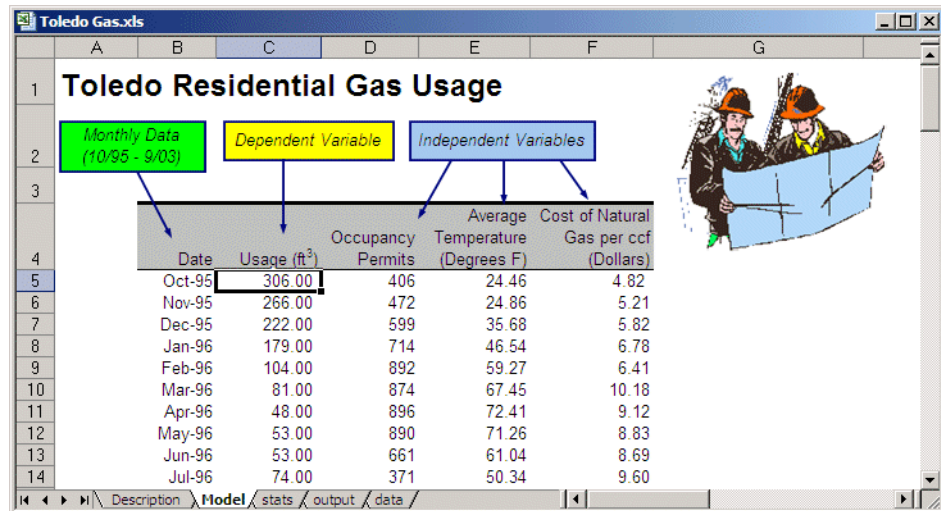


Figure 3.1 Example spreadsheet

Loading and starting CB Predictor

If you have Crystal Ball Professional or Premium Edition, CB Predictor is loaded when you start Crystal Ball.

To start CB Predictor, use one of these:

- In the menubar, choose Run > CB Predictor; or
- Type Alt-r, p.

When you start CB Predictor, the CB Predictor dialog, or wizard, appears as shown in Figure 3.2. It has four tabs, arranged from left to right in the typical order of their use. For descriptions of each setting on each tab, see Appendix A, “The CB Predictor Wizard.”

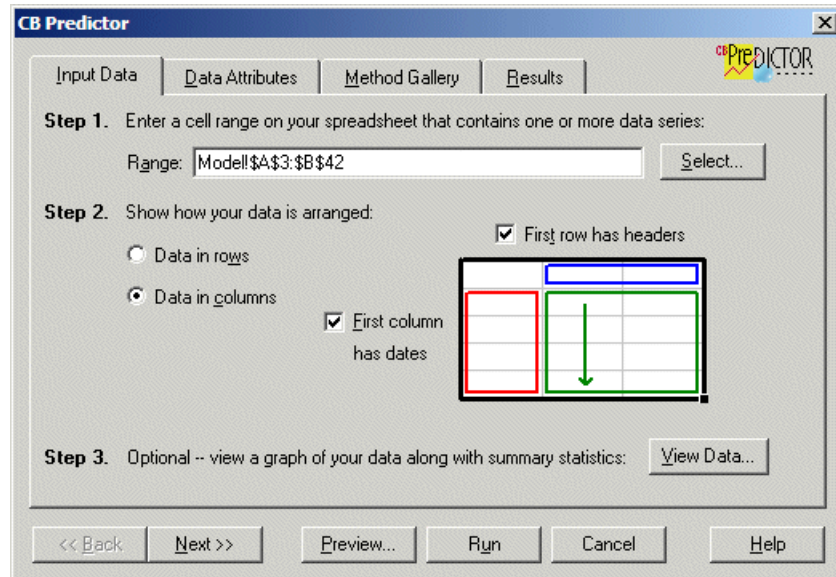


Figure 3.2 The CB Predictor wizard dialog

Guidelines for using CB Predictor

After you create your Excel spreadsheet with historical data, forecasting data using CB Predictor follows a 10-step process:

1. **Select a cell range with your historical data to forecast.**
2. **Specify your data arrangement.**
3. **View a graph of your historical data to identify any seasonality (data cycles) or trend and to see summary statistics.**
4. **Identify your time periods, whether your data have seasonality, and, if so, how long your season is.**
5. **Set whether you want to use multiple linear regression to forecast any variables.**
6. **Select the time-series forecast methods to try for each variable.**
7. **Enter the number of periods you want to forecast.**
8. **Select a confidence interval to calculate or display with your forecasted values.**
9. **Select the results you want.**

10. Preview and run the forecast, creating your results.

CB Predictor leads you through these steps using the CB Predictor wizard, but you can also go directly to any of these steps to change methods or select different settings and reforecast data.

Also, many of these steps might be done automatically by CB Predictor. For example, CB Predictor's Intelligent Input can find and select your data, their arrangement, and the data units. And, you can skip many other steps if the settings are already what you want. For example, all the time-series forecasting methods are selected by default, and that is probably how most users will leave them.

Selecting historical data

When you select your historical data, you must identify the Excel cells that contain the data, including dates and headers, and set settings to identify dates, headers, and orientation in the data.

To produce a reasonable forecast, you should have a minimum number of historical data points. CB Predictor imposes several requirements defining a minimum number of points to use to create a forecast. The bare minimum is 5. However, other limitations include:

- Single moving average requires that the number of historical data points is twice the number of points to forecast.
- Double moving average requires that the number of historical data points be three times the number of points to forecast (or at least 6, whichever is higher).
- To use seasonal methods, you must have at least two seasons (complete cycles) of historical data.
- For multiple linear regression, the number of historical data points must be three times the number of independent variables (counting the included constant as an independent variable).
- To lag an independent variable in multiple linear regression, the number of historical data points must be three times the lag.

If your data has empty cells in the middle of a data series, CB Predictor returns an error. CB Predictor treats zeros in data series as data values. If you are trying to forecast several data series at once, your data series do not have to start at the same time period. However, all the data series must end at the same time period.

When you initially open a spreadsheet, there are three ways to select the historical data to forecast:

- Use CB Predictor’s Intelligent Input.
- Select your data before you start the wizard.
- Select your data after you start the wizard.

Automatic data selection

The easiest way to select data is to select one cell somewhere in your continuous data range before you start the CB Predictor wizard. When you start the wizard, CB Predictor’s Intelligent Input searches for all the adjacent cells with numbers, dates, and headers and makes some other assumptions about your input data, such as whether your data are in rows or columns. This often completes most of the fields and settings on the Input Data tab.

Manual data selection before starting CB Predictor

The second way to select your historical data is to highlight the data range (including headers and dates) before you start CB Predictor.

Manual data selection within CB Predictor

The third way to select historical data is to start CB Predictor with no data, date, or header cells selected. The Input Data tab of the wizard appears with the Range field blank. At this point, you must select your historical data manually.

To select historical data manually from the Input Data tab:

1. Start CB Predictor.

The Input Data tab of the wizard dialog appears. For more information on this dialog, see “Input Data tab” on page 100.

2. Under Step 1, in the Range field, either type a range name (if defined), enter the range of cells with the historical data, including any headers (e.g., A4:B42), or:

a. Click Select.

The Select Range dialog replaces the wizard dialog.

b. Select the cells with the historical data, including any cells with headers and dates.

- c. **Click OK to return to the wizard dialog.**

The selected range appears in the Range field.

Specifying data arrangement

No matter which selection method you use, you must also specify your data arrangement to help CB Predictor identify whether you have dates and headers adjacent to your data series and whether your data are in rows or columns. If you used the Intelligent Input to select your data, these settings should already be set correctly.

To set arrangement settings, use the Input Data tab of the CB Predictor wizard as shown in Figure 3.2 on page 58:

1. **Under Step 2 of the wizard, if your historical data is in:**
 - Rows, select Data In Rows
 - Columns, select Data In Columns
2. **If you have headers (titles) at the top of your columns or to the left of your rows, select the First Row (Column) Has Headers setting.**
3. **If your first column or row lists the dates or time periods for your data series, select the First Column (Row) Has Dates setting.**

Viewing your historical data

As you progress through the wizard, you need to know if your data are seasonal (increase and decrease in a regular cycle) and, if so, what the season or cycle is. If you don't already have a feel for the behavior of your data, you might want to view your selected historical data before you continue.

To view a graph of your historical data:

1. **Under Step 3 of the CB Predictor wizard (on the Input Data tab), click View Data.**

The View Historical Data dialog appears as shown in Figure 3.3.

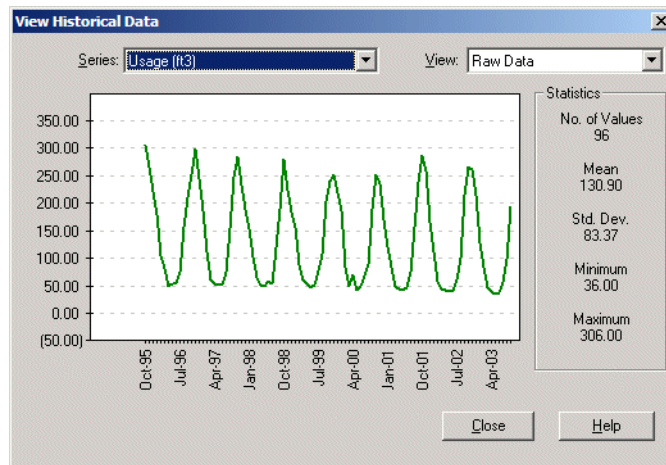


Figure 3.3 View Historical Data dialog — Raw Data view

2. If your data have a recurring pattern, your data might be seasonal, and you should note how many dates or time periods are in one cycle of the pattern.
3. If you selected more than one historical data series, change the graph to view another data series by selecting it from the Series list.

CB Predictor Note: When the Series list is selected, you can also use the up and down arrows to scroll through the list.

4. To see the three highest autocorrelations and the Ljung-Box statistic, select Autocorrelations from the View list.

The View Historical Data dialog appears in Autocorrelations view as shown in Figure 3.4.

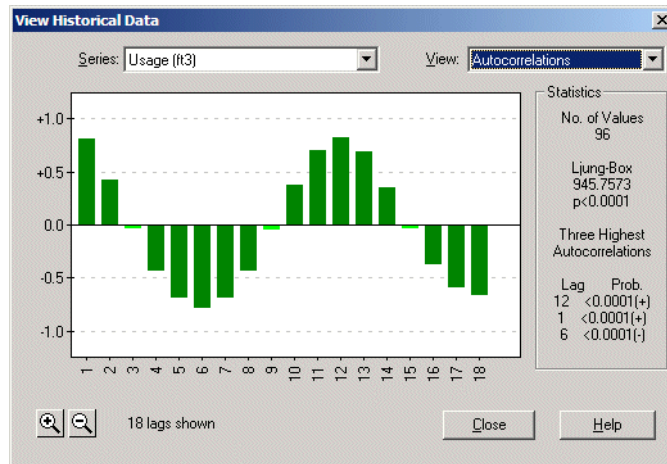


Figure 3.4 View Historical Data dialog — Autocorrelations view

For information on the Ljung-Box statistic, see “Ljung-Box statistic” on page 53.

For more information on both views of the View Historical Data dialog, see page 102.

5. Click Close.

The Input Data tab reappears.

Identifying time periods and seasonality

You need to identify your time periods and seasonality for CB Predictor.

To identify your data’s time periods and seasonality:

1. In the Input Data tab, click Next.

The Data Attributes tab appears. For more information on this tab, see “Data Attributes tab” on page 104.

2. Under Step 4 on the Data Attributes tab, identify the time period for your data values.

For example, if your data represent monthly numbers, select months.

3. Indicate the seasonality of your data:

- If any of your data series are seasonal, select the Seasonality setting and enter the number of time periods it takes before your data

pattern repeats. You must have at least two seasons (complete cycles) of data to use the seasonal methods.

This number is usually the number of periods per year. For example, if you have 24 monthly data points, and your data has peaks every December, your seasonality (repeating pattern) has a period of one year or 12 months.

CB Predictor Note: You can also view the autocorrelations on the View Historical Data dialog to discover how many periods you have in a season.

- If none of your data are seasonal, select the No Seasonality setting.
- If you are forecasting multiple data series and each has a different seasonality, you must forecast each individually.

Using multiple linear regression

Glossary Term:
variables—
In regression, data series are also called variables.

If you know that some *independent variables* affect another variable of interest, you should use multiple linear regression as the forecasting method for that particular *dependent variable*. For example, summer temperatures affect electricity usage because as it gets hotter, more people run their air conditioning. This means that electricity usage (the dependent variable) is dependent on the temperature (an independent variable).

CB Predictor Note: The “multiple” in multiple linear regression represents the fact that you can have more than one independent variable.

To forecast a dependent variable with regression, CB Predictor:

- Creates an equation that defines the mathematical relationship between the independent variables and a dependent variable. This is the regression equation.**
- Forecasts each independent variable by running all the selected time-series forecasting methods for each one and using the best method for each.**
- Calculates the regression equation with the forecasted independent variable values to create the forecast for the dependent variable.**

This process of creating the regression equation, forecasting the independent variables, and calculating the results to forecast the dependent variable *in one easy step* is called HyperCasting™.

To use multiple linear regression:

1. **Under Step 5 on the Data Attributes tab, if one or more variables depend on other variables that you have, select the Use Multiple Linear Regression setting.**

The Regression Variables dialog appears as shown on page 106.

2. **Usually the dependent variable or variables already appear in the Dependent Variables list. If they do not, follow these steps:**
 - a. **Select the name of your dependent variable in the All Series list.**

You can have more than one dependent variable. CB Predictor forecasts them all, one at a time, as functions of all the same independent variables.

- b. **Click >> next to the Dependent Variables list.**

The variable moves to the Dependent Variables list.

3. **Verify that all independent variables are included in the Independent Variables list. If not, add them the same way:**
 - a. **Select the names of your independent variables in the All Series list.**

To select multiple names, hold down either the <Ctrl> key or the <Shift> key or drag the mouse over the list.

- b. **Click >> next to the Independent Variables list.**

The variables move to the Independent Variables list.

4. **To lag independent variable data by a number of time periods:**

- a. **Select a variable from the Independent Variable list.**
 - b. **Enter a number in the Lag field at the bottom of the list.**
 - c. **Repeat for any other independent variables you want to lag.**

5. **For any independent variables you don't want to forecast:**

- a. **Select the variable from the Independent Variable list.**
 - b. **Select the Do Not Forecast setting at the bottom of the list.**
 - c. **Repeat for any other independent variables you don't want CB Predictor to forecast.**

6. Click OK.

The Data Attributes tab reappears.

7. Select the regression method to use, either standard, forward stepwise, or iterative stepwise.

8. If you selected a stepwise regression, you can set settings associated with stepwise regression.

a. Click Stepwise Options.

The Stepwise Options dialog appears as shown on page 108.

b. Set the settings. For more information on these settings, see “Regression methods” on page 48.

c. Click OK.

You return to the wizard.

9. If you want CB Predictor to calculate the regression equation without a constant (to force the resulting equation to pass through the mathematical origin), be sure the Include Constant setting is not selected.

Selecting time-series forecasting methods

You can forecast historical data using many different time-series forecasting methods. Some methods are designed to work best for certain types of data:

- Seasonal data (increasing or decreasing in a regularly recurring pattern over time)
- Trend data (consistently increasing or decreasing over time)

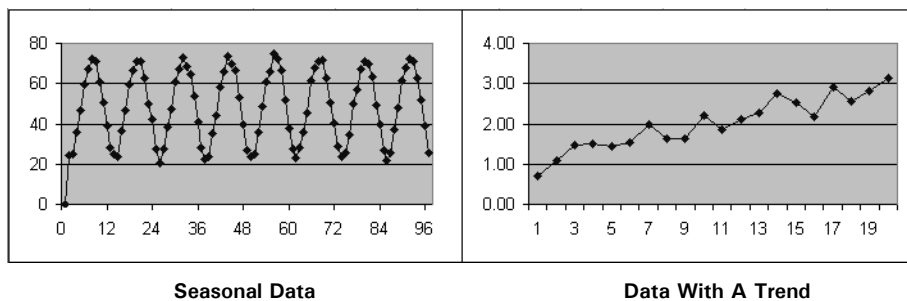


Figure 3.5 Seasonal data and data with a trend

For time-series forecasting, any of the time-series forecasting methods should work with different amounts of success. However, each method has its own “specialty,” as described in the following table.

Table 3.1 Choosing a forecasting method

<i>No trend or seasonality</i>	<i>Trend only</i>	<i>Seasonality only</i>	<i>Both trend and seasonality</i>
Single Exponential Smoothing	Holt’s Double Exponential Smoothing	Seasonal Additive	Holt-Winters’ Additive
Single Moving Average	Double Moving Average	Seasonal Multiplicative	Holt-Winters’ Multiplicative

In addition to this breakdown, there are two types of seasonal methods: additive and multiplicative. Additive seasonality has a steady pattern amplitude, and multiplicative seasonality has the pattern amplitude increasing or decreasing over time.

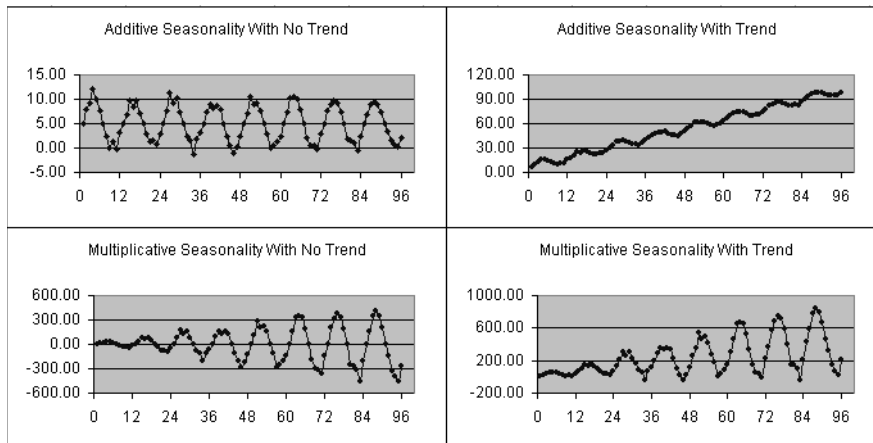


Figure 3.6 Different seasonal curves

You can use a scatter plot of your data (using either Excel or the View Data button on the Input Data tab) to decide whether you have trend or seasonal data. This can help you decide which methods will work best when forecasting the data. However, selecting all the time-series forecasting methods does *not* significantly slow down the calculations unless you are forecasting thousands of values at once, so you might as well try them all.

To select the forecasting methods to use:

1. In the Data Attributes tab, click Next.

The Method Gallery tab appears as shown in Figure 3.7.

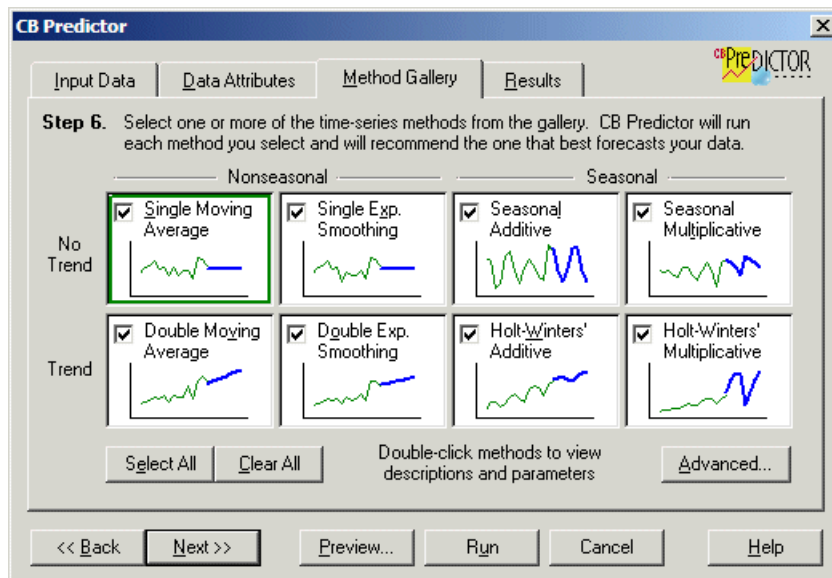


Figure 3.7 Method Gallery tab

For a description of the Method Gallery tab settings, see “Method Gallery tab” beginning on page 110.

2. In the Method Gallery tab, either:

- Try all the methods by clicking on Select All.
- Try only some methods by clicking on Clear All and then clicking in the checkbox for each method you want to try.

3. To manually set the parameters for any method, overriding the automatic calculation of parameters:

a. Double-click in the method area.

The method’s parameter dialog appears, similar to Figure A.8 on page 111.

b. Select the User Defined setting.

The parameter fields become active. You can reset the method to automatically optimize the parameters at any time by selecting the Automatic setting.

c. Enter the parameter values in the parameter fields.

For more information on these parameters, see “Seasonal smoothing parameters” on page 41.

d. Click OK.

The parameters dialog closes.

CB Predictor Note: *The user-defined settings remain until you reset them. A double asterisk next to the method in the Method Gallery indicates that the method is set to use user-defined parameters.*

4. Set advanced settings.

For more information on setting advanced settings, see “Selecting error measures” below and “Selecting forecasting techniques” on page 70.

Selecting error measures

CB Predictor uses one of three error measures to determine which time-series forecasting method works best:

RMSE	Root Mean Squared Error
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error

When determining the best method, CB Predictor calculates the selected error measure when fitting each method to the historical data. The method with the lowest error measure is considered best, and the rest of the methods are ranked accordingly. For more information on these error measures, see “Time-series forecasting error measures” on page 42.

By default, CB Predictor uses RMSE to select the best method.

To change which error measure CB Predictor uses:

1. In the Method Gallery, click Advanced.

The Advanced Options dialog appears, similar to Figure A.9 on page 113.

2. Select the error measure you want CB Predictor to use to determine the best method.

For more information on the Advanced Options dialog, see “Advanced Options dialog” on page 113.

3. Click OK.

The Method Gallery reappears.

Selecting forecasting techniques

CB Predictor uses one of four forecasting techniques for time-series forecasting as shown in the following table:

Table 3.2 Choosing a forecasting technique

<i>Technique</i>	<i>Optimizes the forecasting parameters to minimize the error measure using the technique:</i>
Standard Forecasting	Error measure between the fit values and the historical data for the same period.
Simple Lead	Error measure between the historical data and the fit offset by a specified number of periods (lead).
Weighted Lead	<i>Average</i> error measure between the historical data and the fit offset by 0, 1, 2, etc. periods, up to the specified number of periods (weighted lead).
Holdout	Error measure between a set of excluded data and the forecasting values. CB Predictor does not use the excluded data to calculate the forecasting parameters.

By default, CB Predictor uses standard forecasting to select the best method.

To change which forecasting technique CB Predictor uses:

1. In the Method Gallery, click Advanced.

The Advanced Options dialog appears, similar to Figure A.9 on page 113.

2. Select the forecasting technique you want CB Predictor to use for time-series forecasting.

For more information on the Advanced Options dialog, see “Advanced Options dialog” on page 113.

3. If you selected Simple Lead, Weighted Lead, or Holdout, enter the appropriate lead or holdout in the field by the setting.

4. Click OK.

The Method Gallery reappears.

Entering the number of periods to forecast

After CB Predictor finds the method that best fits the historical data, it is ready to use that same method to forecast future values. You need to decide how many time periods you want to forecast.

As you are trying to decide, there are a few points to keep in mind:

- The first few values are fairly reliable. Only forecast as many values as you need.
- The farther out you try to forecast, the less reliable the forecasted values are. The confidence interval of any forecast grows to reflect this decrease in reliability.
- The maximum number of time periods possible is 100.

To enter the number of periods to forecast:

- 1. In the Method Gallery, click Next.**

The Results tab appears as shown in Figure A.10 on page 115.

- 2. Under Step 7, enter the number of periods you want to forecast.**

For details, see “Results tab” beginning on page 115.

Selecting a confidence interval

The confidence interval defines the range above and below a forecasted value where the value has some probability of occurring. For example, a confidence interval of 10% and 90% gives two points for each forecasted value. The lower point represents the 10th percentile. The higher point represents the 90th percentile. The farther out the forecast is, the larger this range is.

To set a confidence interval, under Step 8 on the Results tab, select the confidence interval you want to calculate and display with your results. For details, see “Results tab” beginning on page 115.

Select your results

After you identify your historical data, select the methods to try, and decide how many time periods you want to forecast, you must decide which results you want to create. CB Predictor can:

- Paste the forecasted data anywhere on your worksheet or into a new worksheet
- Create a chart that can show historical data, fitted values, and forecasted data and its confidence interval
- Generate a report summarizing the findings
- Create a PivotTable of all the historical data, fitted values, forecasted data, and confidence intervals
- Create a PivotTable of some or all the method information for each forecast, including the errors, parameters, and statistics for each method tried

You can choose to generate only basic results, such as pasted data and charts, until you are sure you have the forecasts you want. Then you can create customized, detailed reports and tables to present to others.

To generate the results you want:

1. **Under Step 9 on the Results tab (page 115), to paste forecasted values:**
 - a. **Select the Paste Forecasts setting.**

Enter the cell reference of the starting location in the field to the right. The default starting location pastes the forecasted values immediately at the end of your historical data.
 - b. **To paste the data somewhere other than the end of the data, enter the cell where you want the pasted data to start or click Select to select a different cell interactively.**
 - c. **Select to paste the forecasted values in rows or columns.**
2. **Select the other types of output you want by clicking on the appropriate checkboxes.**

3. To customize the output types:

a. Click Preferences.

The Preferences dialog appears as shown on page 117.

b. Click the tab of the result you want to customize.

c. Change the settings.

For more information on these settings, see “Preferences dialog” on page 117.

d. Repeat steps b and c for any other result you want to customize.

e. Click OK.

The Results tab of the wizard reappears.

4. Enter a title in the Title field to identify your output.

This title appears on all the result sheets that CB Predictor creates. The default title is the worksheet name.

Previewing and running the forecast

CB Predictor actually calculates the forecasts for each data series when you select either Preview or Run. The difference is that Preview lets you view the forecasts and make changes before it generates any of the results. Run simply generates the results.

1. To preview the forecast before generating the selected results, click Preview.

The Preview dialog appears. For more information, see “Preview Forecast dialog” on page 125.

2. To run the forecast and produce the selected results, click Run.

CB Predictor Note: *You can click Preview or Run from any of the wizard tabs at any time, as long as the data have been properly defined in the Input Data tab.*

Analyzing the results

CB Predictor simplifies the actual forecasting process, but you still must understand the results it gives you.

Pasted forecasts

Two settings in the Results tab affect pasted data: the Paste Forecasts setting and the number of periods to forecast. If you selected Paste Forecasts, when you click Run, CB Predictor pastes the number of forecasted values at the end of your historical data or somewhere else if you specify another location. The worksheet automatically scrolls to where the pasted data begin.

The forecasted values can appear in bold or italics, depending on the selected settings in the Paste Preferences dialog.

Even though CB Predictor tried all the methods you selected in the Method Gallery, it generates the pasted values using the best method, unless you overrode the best method. In that case, the program generates the pasted values using the overriding method.

Statistical Note: *Of the eight time-series forecasting methods, two result in flat lines: single moving average and single exponential smoothing. The forecasted values for these are all the same. This isn't an error. It is the best possible forecast for volatile or patternless data.*

When you paste regression results, the independent variable forecast values are pasted as simple value cells. The dependent variable forecasted values are created as formula cells with the regression equation as the formula. The regression equation coefficients appear below the pasted values.

If you did not generate any charts with your results, you can create an Excel chart by selecting the forecasted values (along with some historical data) and using Excel's Chart wizard.

Charts

Three settings in the Results tab affect charts:

Number Of Periods To Forecast	Sets how many forecasted values appear on the chart.
Confidence Interval	Sets which confidence interval to calculate. If you select None, you cannot display a confidence interval on your charts.
Chart	Selects whether to generate charts.

If you select Chart and click Run, CB Predictor generates Excel charts for each series with the number of forecasted values. The charts appear either on a different worksheet of the current workbook or on a different workbook, depending on the selected settings in the Chart Preferences dialog. The data used to generate the charts appears to the right of the charts.

Even though CB Predictor tries all the methods you selected in the Method Gallery, it generates the chart values using the best method, unless you override the best method. In that case, the program generates the chart values using the overriding method.

On the chart, the green line represents your historical data, the blue lines represent both the fitted and forecasted values, and the red lines above and below the forecasted values represent the upper and lower confidence interval. There is a gap between where the historical and fitted values end and the forecasted values begin to delineate the past and future values.

Statistical Note: *Of the eight time-series forecasting methods, four result in straight lines: single moving average, single exponential smoothing, double moving average, and Holt's double exponential smoothing. Only the seasonal methods and multiple linear regression result in curves that try to approximate any repeated data pattern.*

Excel Note: *You can print your charts in black and white from Excel by selecting File > Page Setup > Sheet > Black And White.*

Reports

Four settings in the Results tab affect reports:

Number Of Periods To Forecast	Sets how many forecasted points appear in the report.
Confidence Interval	Sets which confidence interval to calculate. If you select None, you cannot display a confidence interval on your report.
Title	Specifies the title that appears at the top of the report.
Report	Selects whether to generate a report.

If you select Report and click Run, CB Predictor generates one report for your entire forecast. The report appears either on a different worksheet of the current workbook or on a different workbook, depending on the selected settings in the Report Preferences dialog.

Unlike the pasted data and charts, the report can have information about one or all of the methods selected in the Method Gallery. For example, if you selected to include the forecasted values and charts, CB Predictor creates those using the best method. You can include statistics and parameters, however, for each method tried.

Excel Note: You can print your report in black and white from Excel by selecting *File > Page Setup > Sheet > Black And White*.

Results table

Three settings in the Results tab affect the Results table:

Number Of Periods To Forecast	Sets how many forecasted values appear for each method on the Results table.
Confidence Interval	Sets which confidence interval to calculate. If you select None, you cannot display a confidence interval on your Results table.
Results Table	Selects whether to generate a Results table.

If you select Results Table and click Run, CB Predictor generates one Results table for all the series. The Results table appears either on a different

worksheet or on a different workbook, depending on the selected settings in the Results Table Preferences dialog.

Even though CB Predictor tried all the methods you selected in the Method Gallery, it generates the Results table using the best method, unless you overrode the best method. In that case, the program generates the result values using the overriding method.

Methods table

Only one setting in the Results tab affects the Methods table: the Methods Table setting, which selects whether to generate it.

If you select Methods Table and click Run, CB Predictor generates one Methods table for your entire forecast. The Methods table appears either on a different worksheet or on a different workbook, depending on the selected settings in the Methods Table Preferences dialog.

The Methods table reports all the parameters and statistics for one or all the methods you selected in the Method Gallery. In the table, the methods appear in the order from the Method Gallery. The method used to generate the forecasted values, either the best method or the overriding method, is highlighted in bold, blue text. The method is likely to be different for each forecasted series.

To compare the quality of the results of different time-series forecasting methods, check the errors: RMSE, MAD, and MAPE. For all of these, the smaller the better. So, you can compare the RMSE of one method to the RMSE of another method, and the smaller one should be ranked better. However, you cannot compare the RMSE of one method to the MAD or MAPE of another method. For more information, see “Time-series forecasting error measures” on page 42.

To compare the quality of a regression, look for the following values:

Table 3.3 Evaluating regression quality

<i>Statistic</i>	<i>Range</i>	<i>Ideal value</i>	<i>Shows that:</i>
R^2 or Adjusted R^2	0 to 1	Close to 1	The linear regression accounts for almost all the variability in the dependent data.
F probability	0 to 1	Less than 0.05	The quality of the overall regression (dependency of the dependent variable on the independent variables) is good.
t probability	0 to 1	Less than 0.05	The quality of the coefficient of the regression equation is good.
Durbin-Watson	0 to 4	2	No auto-correlation (at lag 1) exists.
Theil's U	Greater than 0	Less than 1	The quality of the results is better than guessing.

For more information on these statistics, see “Regression statistics” on page 146.

Customizing reports and charts

You can customize reports and charts by customizing the Excel templates that control how they look. The CB Predictor folder contains these two template files: Report and Charts.

To customize reports or charts:

1. **In the CB Predictor folder, open the appropriate template file.**
2. **Make any style changes you want.**

For example, in the report, you can change the font of the title, the width of the columns, or turn on grid lines. For the chart, you can change the chart type or the color of the upper confidence interval and fitted lines, change the background color, or change the angle of the horizontal axis labels.

3. **Save and close the template file.**

The next time CB Predictor generates a report or chart, it uses the modified template and appears with the new style.

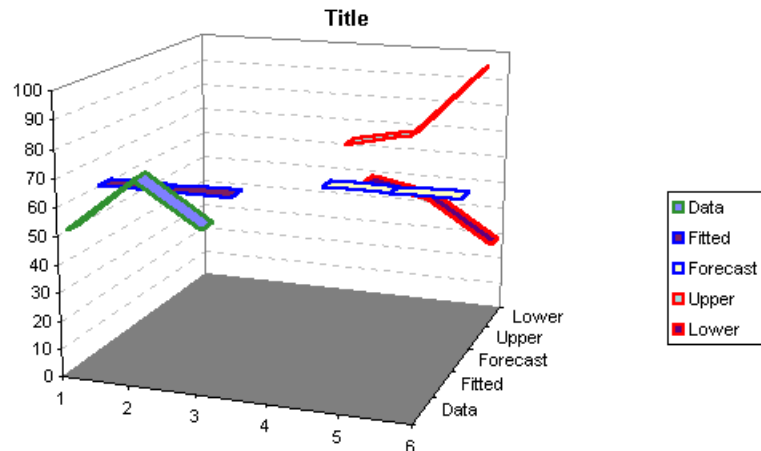


Figure 3.8 Chart as a 3D ribbon chart



Chapter 4

Examples

In this chapter

- Overview
- Inventory control
- Company finances
- Human resources

This chapter describes a detailed Excel model for Monica's Bakery. The workbook has worksheets for sales data, operations, and cash flow. The Sales Data worksheet contains all the historical sales data that is available for forecasting. The Operations worksheet calculates the amount of different ingredients required to make different quantities of the three different breads. The Cash Flow worksheet calculates how much money the bakery has to spend on various capital projects. The Labor Costs worksheet estimates the increase in hourly wages to decide whether to invest in labor-saving equipment.

Examples show to use this type of information to appropriately stock bread ingredients, plan for major purchases such as constructing a silo or buying a delivery van, and estimate the value of the business.

Overview

Monica's Bakery is a rapidly growing bakery in Albuquerque, New Mexico. It opened in March of 2000, and Monica has kept careful records (in an Excel workbook) of the sales of her three main products: French bread, Italian bread, and pizza. With these records, she can better predict her sales, control her inventory, market her products, and make strategic, long-term decisions. To see Monica's workbook, in the CB Predictor Examples folder, open *Bakery.xls*. By default, the file is stored in this folder: C:\Program Files\Decisioneering\Crystal Ball 7\Examples\CB Predictor Examples.

These numbers are ready to forecast. The following examples track Monica's decision-making processes as she works through both short-term and long-term decisions.

Inventory control

The initial reason that Monica needs to forecast is to maintain enough ingredients to keep up with her production. Monica's Bakery places regular orders for ingredients, and Monica's distributors give her discounts for buying in bulk. However, she must balance this savings with the freshness of her products, which requires using the freshest ingredients possible.

In the past Monica scheduled deliveries to her business of bagged flour and other ingredients whenever "it looked low." Sometimes this required paying for express delivery when demand was high or, when the demand after placing the order was unexpectedly low, letting ingredients sit unused until they were no longer fresh. With better forecasting, Monica wants to place orders that give her the best buying power while maintaining the quality of her products.

The Sales Data worksheet shows the daily sales data of each of these products from the opening until the end of June 2002.

Date	French Bread	Italian Bread	Pizza	Total
5-Mar-00	\$591.64	\$852.96	\$187.04	\$1,631.64
6-Mar-00	582.36	1101.63	187.29	1871.28
7-Mar-00	591.63	1008.88	186.79	1787.31
8-Mar-00	589.60	1071.82	186.72	1848.14
9-Mar-00	582.48	1038.78	186.88	1808.14
10-Mar-00	579.34	1002.81	186.95	1769.10
11-Mar-00	606.50	995.83	187.13	1789.46
12-Mar-00	611.57	837.60	187.07	1636.25
13-Mar-00	605.11	1125.16	187.29	1917.55
14-Mar-00	617.31	1053.87	186.93	1858.11
15-Mar-00	630.87	1143.00	187.50	1961.37
16-Mar-00	656.43	1049.61	187.04	1893.08
17-Mar-00	667.51	1271.85	187.42	2126.78
18-Mar-00	658.08	1140.49	187.08	1985.65
19-Mar-00	670.71	1174.30	187.29	2032.30

Figure 4.1 Bakery Sales Data worksheet

Monica created a PivotTable in Excel that summarizes the data for her three main products by week at the bottom of the Operations worksheet. By creating a PivotTable from the data, Monica can change the table to summarize her results by product, by time period, and more.

Demand	Sales Forecast	Selling Price	Units	Unit Weight	Total Weight
French Bread	\$ -	\$ 1.24	-	0.25	-
Italian Bread	\$ -	\$ 2.08	-	1.05	-
Pizza	\$ -	\$ 14.71	-	3.12	-

Ingredients	Forecast Weight	Per Pound	Pounds Required
French Bread	0		
Flour		0.32	0
Yeast		0.0010	0
Salt		0.0015	0

Figure 4.2 Bakery Operations worksheet

Monica wants to order monthly, one month in advance. The bakery has already received this month's delivery, which she placed last month. This month, she must place the order that will be delivered at the end of this month for the next month, so she must forecast sales for the next two months. Since she is in week 173 of her business, the forecast is for weeks 174 to 181.

To forecast the sales for weeks 174 to 181:

1. In the Bakery.xls workbook, click the Operations tab.

The Operations worksheet appears.

2. Select one cell in the Historical Demand By Week PivotTable at the bottom of the worksheet.

3. Start CB Predictor.

CB Predictor's Intelligent Input automatically selects all the PivotTable data.

4. Make sure:

- The cell range is selected correctly, with headers, dates, and data in columns (Steps 1 and 2)
- The time periods are in weeks with a seasonality of 52 weeks (Step 4)
- Multiple Linear Regression is off (Step 5)
- All time-series methods are selected (Step 6)
- The number of periods to forecast is 8 (Step 7)
- The only result selected is Paste Forecasts at the bottom of the PivotTable (Step 9)

5. Click Run.

The results paste to the end of the PivotTable as shown in Figure 4.3.

	B	C	D	E	F
210	170	17,296.23	10,152.48	1,471.40	
211	171	17,541.78	9,657.03	1,488.27	
212	172	17,634.05	9,142.77	1,490.36	
213	173	17,380.54	8,758.17	1,483.24	
214	174	17,366.86	8,599.04	1,476.13	
215	175	17,352.91	8,385.96	1,469.01	
216	176	17,338.95	8,463.82	1,461.88	
217	177	17,325.00	8,450.91	1,454.76	
218	178	17,311.04	8,367.91	1,447.64	
219	179	17,297.09	8,305.56	1,440.52	
220	180	17,283.13	8,813.38	1,433.40	
221	181	17,269.18	9,055.13	1,426.28	

Figure 4.3 Forecasted bakery operations results

The last four weeks of forecast values for each data series are automatically summed and placed into the table at the top of the spreadsheet, in the Sales Forecast column (cells C9:C11). In this table, the monthly sales forecast is converted to the number of items sold and then into the weight of each product.

The second table (below this top table) takes the total weight of each product (in cells C15, C19, C23) and calculates how much of each ingredient is required to produce that much product. The ingredients for each are then summed in the third table (below the second table) into the total amounts to order for the month (cells D31 to D34).

	A	B	C	D	E	F	G	H
26		Salt		0.00	1			
27		Tomato		0.10	122			
28								
29								
30		Consolidated Requirements	Pounds					
31		Flour		11,235				
32		Yeast		52				
33		Salt		39				
34		Cheese		122				
35								
36								
37								

Figure 4.4 Ingredient forecast for the last four weeks

Based on the forecast, Monica should order:

- 11,235 pounds of flour
- 52 pounds of yeast
- 39 pounds of salt
- 122 pounds of cheese

Company finances

Monica is always concerned about the bakery's month-to-month cash flow (on a percent of sales basis). Not only can CB Predictor help her manage her inventory, she can use it to predict her revenue and understand her cash flow situation better. Understanding the bakery's cash flow can, in turn, help her time major capital expenditures better.

There are two major capital expenditures Monica is considering for the bakery: a flour silo and a delivery van. She wants to start construction on the silo in July and purchase the new delivery van in August. She needs to forecast when the bakery can safely pay for these projects or whether the bakery must finance them.

The bakery cash flow information is laid out on the Cash Flow worksheet, shown in the next figure.

	Common-Sized	July	August	September
Revenue Forecast	100%	\$ -	\$ -	\$ -

Expenses	Common-Sized	July	August	September
Cost of Goods				
Fixed	6.708	6,708	6,708	6,708
Variable	23%	-	-	-
Overhead				
Fixed	8.924	8,924	8,924	8,924
Variable	18%	-	-	-
Financing	5%	-	-	-
Taxes	17%	-	-	-
Total Expenses		\$ 15,632	\$ 15,632	\$ 15,632

Figure 4.5 Bakery Cash Flow worksheet

This worksheet has a PivotTable at the bottom that summarizes the sales data for the bakery's three main products by month. You can forecast the next three months of revenue to decide when to attempt the capital expenditures.

To forecast the next three months of revenue:

1. In the Bakery.xls workbook, click the Cash Flow tab.

The Cash Flow worksheet appears.

2. Select one cell in the Historical Revenue By Month PivotTable at the bottom of the worksheet.

3. Start CB Predictor.

CB Predictor's Intelligent Input automatically selects all the PivotTable data.

4. Make sure:

- The cell range is selected correctly, with data in rows and no dates or headers (Steps 1 and 2)
- The time periods are in months with a seasonality of 12 months (Step 4)
- All time-series methods are selected (Step 6)
- The number of periods to forecast is 3 (Step 7)
- The only result selected is Paste Forecasts At Cell \$AQ\$35 (Step 9)
- The pasted results are set to Paste In Rows

5. Click Run.

The results paste into the table at the bottom of the worksheet, cells AQ36 to AS36 and also appear at the top of the worksheet (cells E4 to G4) as shown in Figure 4.6 on page 88.

The revenue forecasts for the next three months are used to calculate the percentage expenses in the second table.

The second table calculates the total expenses, and the third table calculates the necessary expenditure for each extraordinary item. Below these tables is the cash flow summary for the next three months, based on the forecasts. The net cash at the end of each month is what Monica is looking for.

Bakery.xls										
	A	B	C	D	E	F	G	H	I	J
1	Cash Flow Analysis (Percent of Sales Basis) for Monica's Bakery									
2										
3				Common-Sized	July	August	September			
4	Revenue Forecast		100%	\$124,351.87	\$125,870.54	\$127,389.21				
5										
6										
7				Common-Sized	July	August	September			
8	Expenses									
9	Cost of Goods									
10	Fixed		6.708%	6,708	6,708	6,708				
11	Variable		23%	28,601	28,950	29,300				
12	Overhead									
13	Fixed		8.924%	8,924	8,924	8,924				
14	Variable		18%	22,383	22,657	22,930				
15	Financing		5%	6,218	6,294	6,369				
16	Taxes		17%	21,475	21,737	21,999				
17	Total Expenses			\$ 94,308	\$ 95,269	\$ 96,230				
18										
19	Extraordinary Items									
20	Silo Construction			50,000.00	-	-				
21	New Van			-	35,000.00	-				
22										
23										
24	Monthly Cash Flow			\$ (19,956)	\$ (4,399)	\$ 31,159				
25										
26	Net Cash at Beginning of Month			\$ 42,941	\$ 22,985	\$ 18,586				
27	Net Cash at End of Month			\$ 22,985	\$ 18,586	\$ 49,745				
28										
29	Minimum Cash Target			20,000	20,000	20,000				
30										

Figure 4.6 Monthly net cash flow results

Based on the forecast, the bakery might be better off waiting another month, until September, before they try to purchase the van.

Human resources

Monica's Bakery is a labor-intensive operation that pays a very competitive wage. However, to maintain her target profitability, Monica must control her labor costs. She knows there are many things done around the bakery that could be done by expensive machinery, such as kneading, mixing, and forming. By accurately predicting her labor costs, she can decide when to invest in some of this equipment to keep her total expenses within budget.

From her interest in economics, Monica knows that a few key macro-economic figures drive labor costs, such as the Industrial Production Index, local CPI, and local unemployment. All of these figures are available on the Internet on

a monthly basis from the Bureau of Labor Statistics and the Department of Commerce.

Monica has created her Labor Costs worksheet with a PivotTable at the bottom that lists her average hourly wage for each month and the monthly numbers for these three indicators.

Labor Cost Analysis for Monica's Bakery

	Jun-03	Dec-03
Wages	\$ 14.27	\$ -
Overhead	33%	33%
Total Employees	18	18
Total Labor Costs	\$ 54,652.74	\$ -
Labor Cost Change		-100%

Average labor costs from regression analysis

Economic Variables for Regression Analysis:

Date	Monica's Average Wage	Industrial Prod. Index	Local CPI	Local Unemployment
Jun-00	\$13.22	121.9	469.0	8.0
Jul-00	13.19	122.6	469.3	8.4
Aug-00	13.24	121.9	470.7	8.2
Sep-00	13.25	121.1	471.7	8.1

Figure 4.7 Bakery Labor Costs worksheet

The average hourly wage depends on or is affected by the other three variables. Because of the dependency, Monica decides to use regression instead of time-series forecasting. For regression, the dependent variable is Monica's Average Wage, and the other three are the independent variables.

To forecast the hourly wage using regression:

1. In the Bakery.xls workbook, click the Labor Costs tab.

The Labor Costs worksheet appears.

2. Select one cell in the Economic Variables for Regression Analysis table at the top of the worksheet.
3. Start CB Predictor.

CB Predictor's Intelligent Input automatically selects all the table data.

4. Make sure:

- The cell range is selected correctly, with headers, dates, and data in columns (Steps 1 and 2)
- The time periods are in months with a seasonality of 12 months (Step 4)
- Multiple Linear Regression is on (Step 5)
- The regression variables are defined: Monica's Average Wage is a dependent variable, all the others are independent variables (Step 5)
- The regression method is set to Standard (Step 5)
- All time-series methods are selected (Step 6)
- The number of periods to forecast is 6 (Step 7)
- The only result selected is Paste Forecast (Step 9)

5. Click Run.

The results paste at the bottom of the table (cells B51 to F56) as shown in Figure 4.8. Notice the results are defined as assumptions.

	B	C	D	E	F
47	Mar-03	13.71	122.7	500.0	9.0
48	Apr-03	14.19	127.3	500.1	8.1
49	May-03	14.12	128.1	500.1	7.6
50	Jun-03	14.27	127.9	500.5	7.4
51	Jul-03	\$14.26	127.9	501.1	7.5
52	Aug-03	\$14.35	127.9	501.7	7.1
53	Sep-03	\$14.41	127.9	502.4	6.8
54	Oct-03	\$14.40	127.9	503.0	6.9
55	Nov-03	\$14.47	127.9	503.6	6.6
56	Dec-03	\$14.54	127.9	504.2	6.3
58	Coefficients for Monica's Average Wage	\$0.00	0.1	0.0	(0.2)

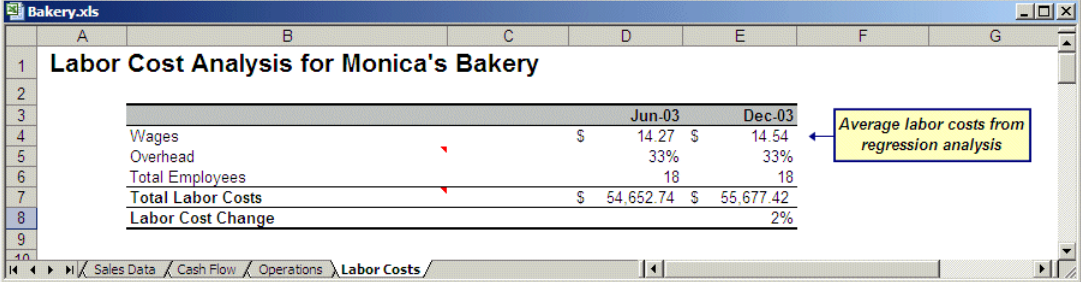
Figure 4.8 Forecasted labor costs for Monica's Bakery

CB Predictor's HyperCasting technology first generates a regression equation to define the relationship between the dependent and independent variables. Second, it uses the time-series forecasting methods to forecast the independent variables individually. Third, CB Predictor uses those forecasted values to calculate the dependent variable values using the regression equation.

The forecast cells of the independent variables are simple value cells. The forecast cells of the dependent variable are formula cells containing the

regression equation and using the other values from the independent variables.

The average wage in December is used to calculate the total increase in her payroll. The increase is only 2%. With these results, Monica decides that labor costs will not increase enough over the next six months to justify a major equipment capital purchase.



	Jun-03	Dec-03
Wages	\$ 14.27	\$ 14.54
Overhead	33%	33%
Total Employees	18	18
Total Labor Costs	\$ 54,652.74	\$ 55,677.42
Labor Cost Change		2%

Average labor costs from regression analysis

Figure 4.9 Predicted labor cost increases

6. Exit the Bakery.xls workbook without saving your changes.

CB Predictor Note: If you save your changes, you will overwrite the example spreadsheet.



Chapter 5

Using CB Predictor with Crystal Ball

In this chapter

- Using CB Predictor with Crystal Ball
- Monica's Bakery example revisited

This chapter describes how to run Crystal Ball with CB Predictor to add risk analysis to your forecast.

It also gives a detailed example of a spreadsheet where you might want to use both forecasting and simulation.

Using CB Predictor with Crystal Ball

"It is far better to foresee even with uncertainty than not to foresee at all."

- Henri Poincaré

in Foundations Of Science

If you are a sales manager, you might be asked to predict future sales to let the manufacturing organization produce sufficient quantities of product to meet demand. However, as shown by the confidence interval around your forecasted data, there is usually a large degree of uncertainty associated with the true future values. How do you account for the uncertainty?

Glossary Term:
assumptions—
Estimated values in a
spreadsheet model that
Crystal Ball defines with
a probability distribution.

Glossary Term:
Crystal Ball forecast—
A statistical summary of
the assumptions in a
spreadsheet model,
output graphically or
numerically.

If you have Decisioneering's Crystal Ball along with CB Predictor, you can define the forecasted values as *assumptions*, create a *Crystal Ball forecast* that represents the entire range of possible outcomes for the forecasted sales, and run a simulation to analyze the effects of the uncertainty. Crystal Ball will generate a forecast chart showing the total amount of product that manufacturing must produce, to meet demand within a specific certainty. You can use the mean of the Crystal Ball forecast or perhaps give manufacturing a range and say that you are 90% certain that the actual number falls in that range. This is the power that combining the two products gives you.

To use CB Predictor and Crystal Ball together, you must understand how to:

- Set CB Predictor to create Crystal Ball assumptions
- Follow a forecast with a simulation

Automatically generating Crystal Ball assumptions

Integrating CB Predictor and Crystal Ball is as easy as choosing a setting that automatically generates forecasted values as Crystal Ball assumptions. To make this setting:

1. On the Results tab, click Preferences.

The Preferences dialog appears.

2. Click the Paste tab.

3. Choose the Paste Forecasts As Crystal Ball Assumptions setting.

4. Click OK.

- 5. Make sure you selected the Paste Forecasts At Cell result.**
- 6. Run your forecast.**

As long as you have Crystal Ball loaded, this setting creates Crystal Ball assumptions as it pastes the forecasted values to your spreadsheet.

The pasted values have a style based on the Style Of Pasted Results settings (bold, by default). The cell color and pattern changes to match the assumption cell preferences set in Crystal Ball under Define > Cell Preferences (green, by default).

Time-series forecast method results

For data series forecasted using time-series methods, CB Predictor creates the assumptions as normal distributions with:

- A mean equal to the forecasted value in the cell
- For time-series forecasted data, a standard deviation calculated using the RMSE

Multiple linear regression results

For multiple linear regression, CB Predictor creates assumptions for the independent variable forecast values only. This is because the independent variable values are simple value cells, but the dependent variable values are formula cells that are a function of the independent variables.

If you want to see the variability of the dependent variable, you can select the pasted formula cells and define them as Crystal Ball forecast cells (by selecting Define > Define Forecast). More likely, you might want to create one formula cell that represents the sum of the data in the dependent variable cells and define that formula cell as a Crystal Ball forecast.

Monica's Bakery example revisited

Monica decides to incorporate uncertainty into her cash flow analysis. For more information on this example, see “Company finances” on page 86.

She had already forecasted her revenue for the next few months, which indicated that she could start construction on her silo in July, but should probably wait to buy her delivery van in August. She has already committed to the silo construction. However, she realizes that her forecasts are point estimates, and she wants to be at least 90% sure that she won't run out of cash and won't be able to meet payroll or other critical bills.

To get a picture of the uncertainty in the bakery's cash flow:

1. In Crystal Ball, select Run > Run Preferences.

The Crystal Ball Run Preferences dialog appears.

2. Set the following settings:

- Run 500 trials
- Use Latin Hypercube sampling
- Use an initial seed value of 999
- Use a sample size of 500

3. Click OK.

4. Open the Bakery.xls workbook.

By default, the file is stored in this folder: C:\Program Files\Decisioneering\Crystal Ball 7\Examples\CB Predictor Examples.

You can also find this example by selecting Help > Crystal Ball > Crystal Ball Examples and selecting the file from the Examples Guide.

5. Click the Cash Flow tab.

The Cash Flow worksheet appears.

6. Select one cell in the Historical Revenue By Month PivotTable at the bottom of the worksheet.

7. Start CB Predictor.

CB Predictor's Intelligent Input automatically selects all the PivotTable data.

8. Make sure:

- The cell range is selected correctly, data in rows, and *no* dates or headers (Steps 1 and 2)
- The time periods are in months with a seasonality of 12 months (Step 4)
- Turn off regression (Step 5)
- All time-series methods are selected (Step 6)
- The number of periods to forecast is 3 (Step 7)
- The only result selected is Paste Forecast at \$AQ\$35 (Step 9)
- In the Preferences dialog, under the Paste tab, the Paste Forecasts As Crystal Ball Assumptions setting is checked (Step 9)

9. Click Run.

The results paste at the bottom of the PivotTable as Crystal Ball assumptions.

10. Define cells E27:G27 as forecasts for Month 1, Month 2, and Month 3, respectively, by selecting each cell and selecting Define > Define Forecast.

11. Run a simulation of 500 trials.

For more information on running a Crystal Ball simulation, see your *Crystal Ball User Manual*.

Monica can now look at the net cash left at the end of each month as a forecast, shown in Figure 5.1.

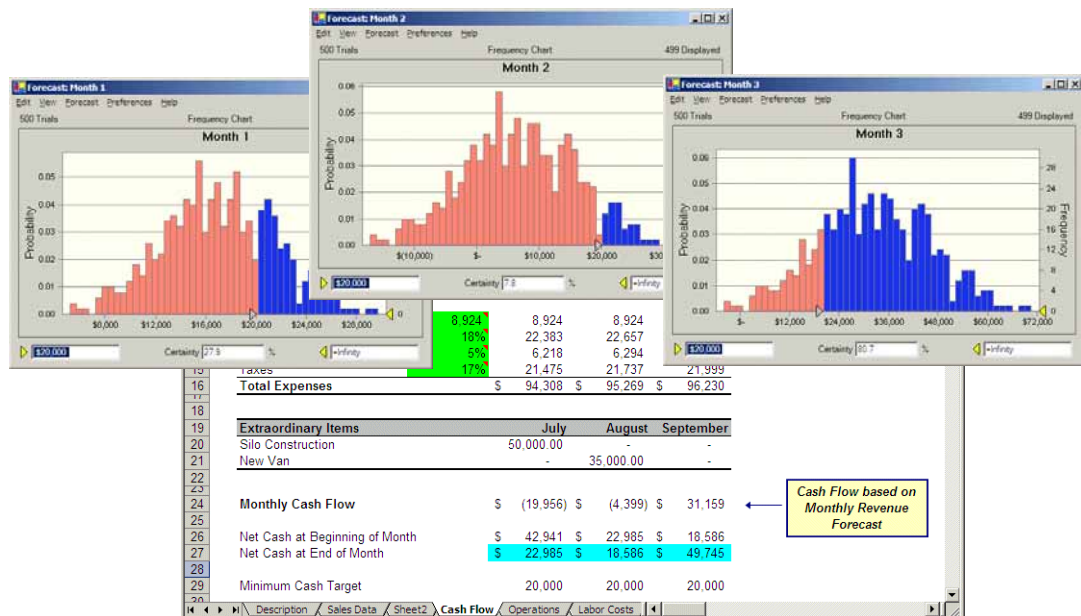
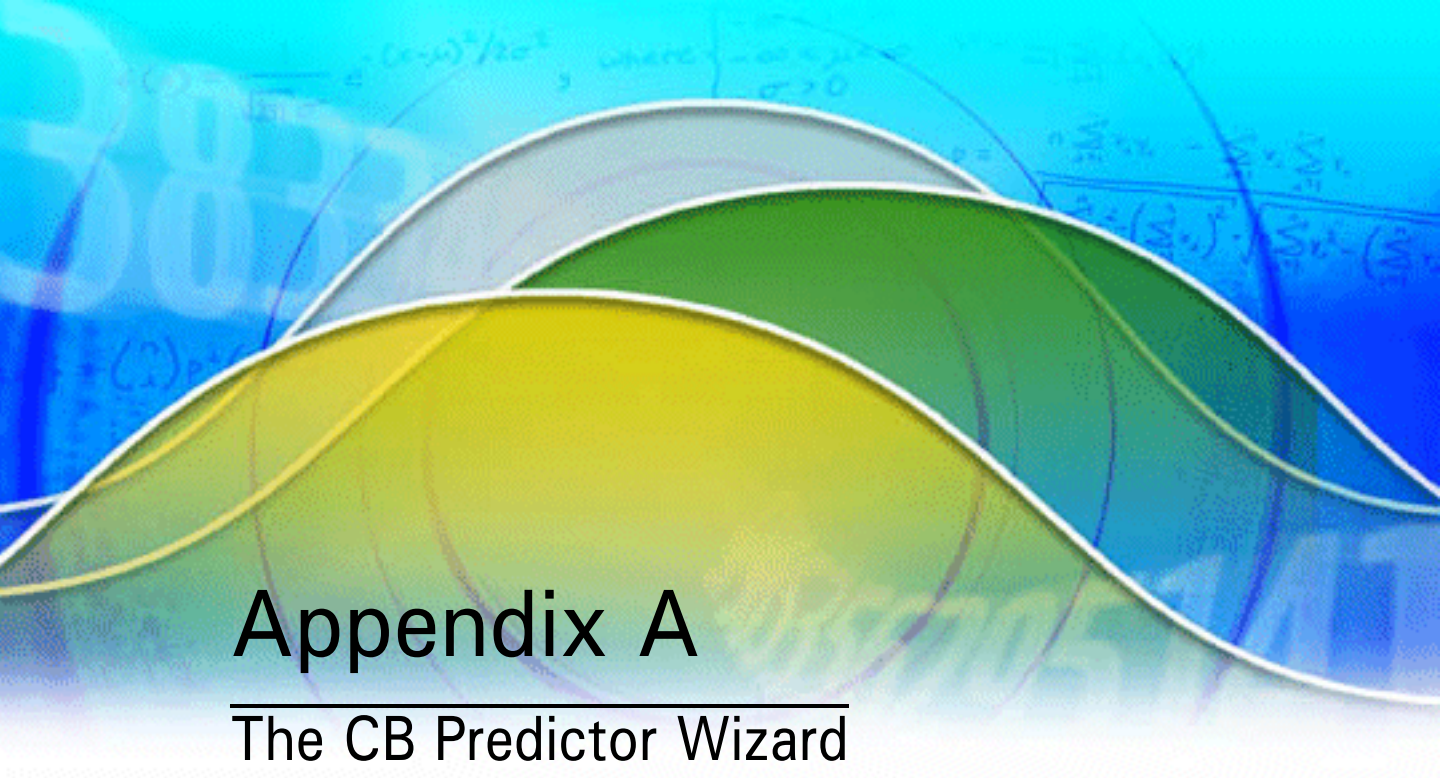


Figure 5.1 Monthly net cash flow results

To be certain that the bakery will not drop below the cash reserve at the end of any month, she slides the left certainty grabber to \$20,000.

Based on the probability of dropping below the reserve threshold, she decides to delay buying the van until September.



Appendix A

The CB Predictor Wizard

In this appendix

This appendix describes the CB Predictor wizard dialog and the settings on each tab:

- Input Data tab
- Data Attributes tab
- Method Gallery tab
- Results tab

When you start CB Predictor, the CB Predictor dialog, or wizard, appears as shown in Figure A.1.

The wizard has four tabs, arranged in typical workflow order:

- Input Data
- Data Attributes
- Method Gallery
- Results

Input Data tab

The Input Data tab of the wizard dialog lets you define your input data.

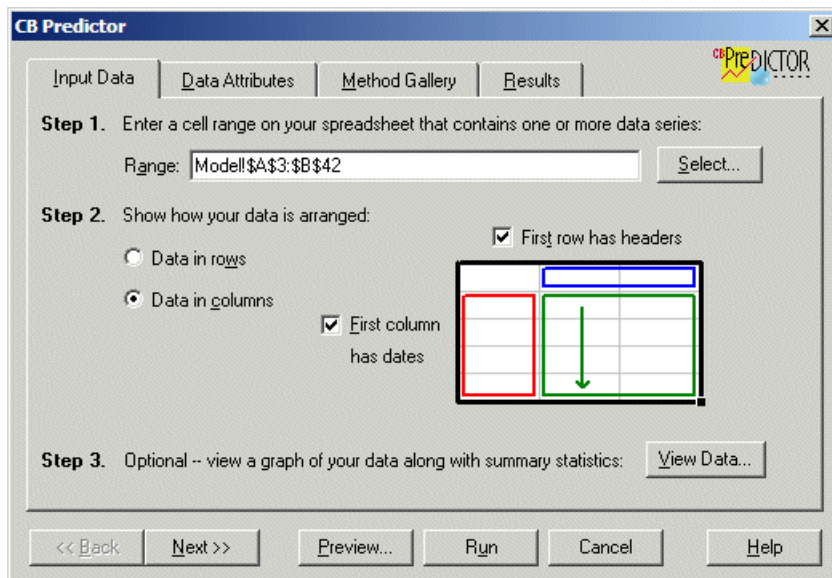


Figure A.1 Input Data tab of the wizard dialog

The fields and buttons are:

Range	<p>Lists the Excel cell range that contains the historical data to forecast, any date fields, and any header fields.</p> <p>If you select one cell before you start the wizard, CB Predictor's Intelligent Input automatically guesses the input range based on the continuously filled cells around the selected cell. If you select a range of cells before you start the wizard, that range appears in the Range field. If you do not select a cell or if you select an empty cell before you start the wizard, you must select your range using the Select button.</p>
Select	<p>Hides the wizard dialog and lets you select a cell range for the Range field. Use this button to select a cell range different from the one in the Range field.</p> <p>This button brings up a small Select Range dialog to let you select the cells directly from the spreadsheet. After selecting the cells, click OK to return to the Input Data tab.</p>
Data In Rows	<p>Indicates that your data are in rows. If you select one cell in your data range before you start the wizard, CB Predictor's Intelligent Input automatically guesses that your data are in the direction with more cells. If the selected cell when you start the wizard is blank, this is the default.</p>
Data In Columns	<p>Indicates that your data are in columns. If you select one cell in your data range before you start the wizard, CB Predictor's Intelligent Input automatically guesses that your data is in the direction with more cells.</p>
First Row (Column) Has Headers	<p>Indicates whether your data have a title or header cell at the top of each column (if your data are in columns) or to the left of each row (if your data are in rows). If you select one cell in your data range before you start the wizard and CB Predictor's Intelligent Input finds text cells at the heads of your data, it automatically selects this setting.</p>
First Column (Row) Has Dates	<p>Indicates whether your data have a first row or column for dates. CB Predictor only recognizes dates in cells that are formatted as Excel dates. If you select one cell in your data range before you start the wizard and CB Predictor's Intelligent Input finds formatted date cells in the first row or column, it automatically selects this setting.</p>
View Data	<p>Opens the View Historical Data dialog and plots your historical data. You can use this to estimate whether your data have any trend or seasonality and what your seasonal period is.</p>

The following buttons appear on all the wizard dialogs:

Preview	Opens the Preview dialog. For more information on this dialog, see “Preview Forecast dialog” on page 125.
Run	Forecasts the selected historical data. The program forecasts the data and provides input using all the methods and settings already selected.

View Historical Data dialog

The View Historical Data dialog lets you view your data to detect the behavior and any seasonality. To display this dialog, click View Data at the bottom of the Input Data tab (Figure A.1, page 100). There are two very different views in this dialog: Raw Data and Autocorrelations.

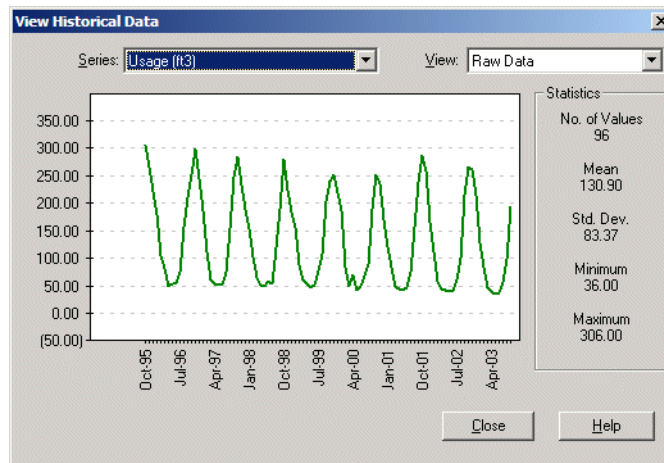


Figure A.2 View Historical Data dialog — Raw Data view

The fields and buttons on this dialog for the Raw Data view are:

Series	Lists all the data series in your spreadsheet. The selected series appears in the chart.
View	Selects whether to display raw data or autocorrelation analysis data in the chart, each with corresponding statistics in the Statistics field.
Statistics	Lists the number of values, mean, standard deviation, minimum, and maximum values for the selected series.

Chart Displays the historical data for the selected series.

View Historical Data — Autocorrelations view

The Autocorrelations view of this dialog lets you display autocorrelations to detect whether the data have any seasonality. The three most significant autocorrelation coefficients (up to a lag of one third of the number of data points) appear in the Statistics area.

The lag represents the number of data periods that the data are offset with the original data before calculating the correlation coefficient. For example, a lag of 12 corresponds with correlating the data with itself, offset by 12 periods; in other words, the correlation of the first data item with the thirteenth data item, the second data item with the fourteenth data item, etc. The value of Prob indicates the significance of the lag.

If the probability of the Ljung-Box statistic is less than 0.05, the set of autocorrelations is significant, and your data are probably seasonal. The seasonality is indicated by the autocorrelation lag. For example, if one of the top three lags is 12 and has a probability of less than 0.001, your data probably have a seasonality of 12 periods.

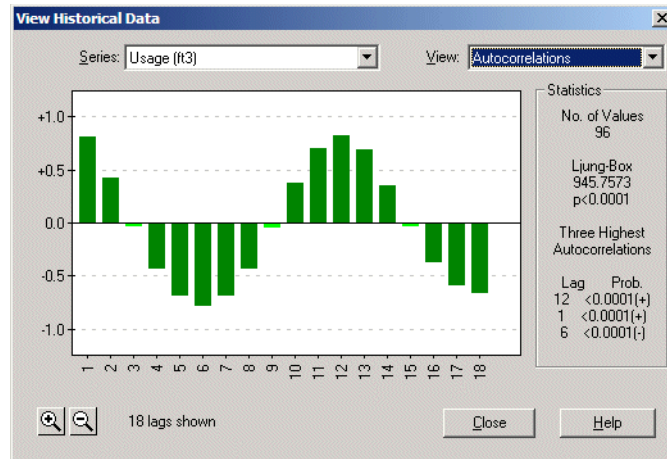




Figure A.3 View Historical Data dialog — Autocorrelations view

The fields and buttons on the View Historical Data dialog for the Autocorrelations view are:

Series	Lists all the data series in your spreadsheet. The selected series appears in the chart.
View	Selects whether to display raw data or autocorrelation analysis data in the chart, each with corresponding statistics in the Statistics field.
Statistics	Lists the number of values, the Ljung-Box statistic, and the top three autocorrelations detected for the selected series, with the lags and associated probabilities.
Chart	Displays the autocorrelation coefficients at different lags for the selected series.
 Zoom In	Increases the magnification of the data graph, showing fewer data points. Zooming is available only in the Autocorrelation view.
 Zoom Out	Decreases the magnification of the data graph, showing more data points. Zooming is available only in the Autocorrelation view.

Data Attributes tab

The Data Attributes tab lets you specify the seasonality of your historical data and use multiple linear regression to forecast any variables that depend on other variable data you might have.

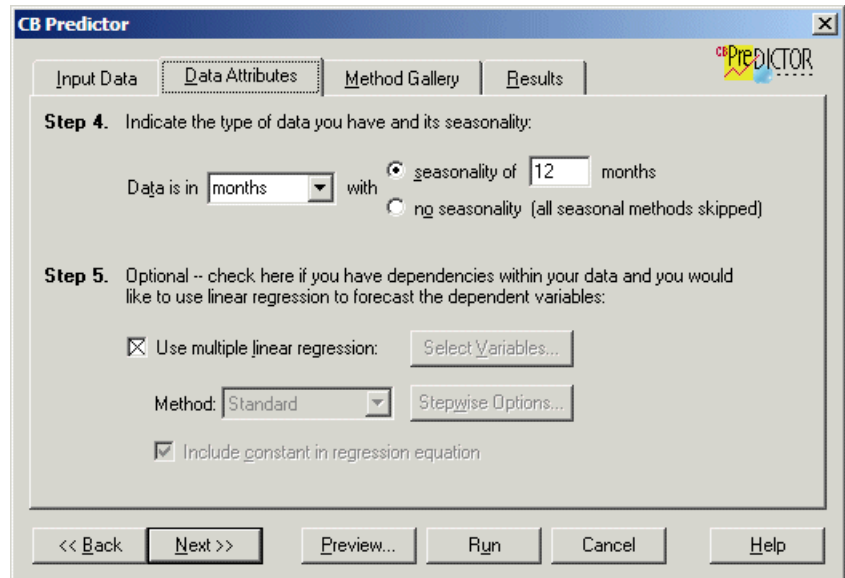


Figure A.4 Data Attributes tab of the wizard

The fields, settings, and buttons in this tab are:

Data Is In	Lets you select what kind of time periods your data points represent. For example, your data might be in hours, days, weeks, months, or quarters. If your data have a time period that is not listed, you can type in your own, and CB Predictor does the rest. The default is Periods.
Seasonality	Indicates that your data are seasonal and sets the number of data points that are in a season (the period of time it takes your data pattern to repeat). For example, if you indicated that your seasonal sales data are in months and you think that your sales pattern repeats every year, you should indicate that you have a seasonality of 12 months. CB Predictor enters a default value in this field based on the time period listed in the Data Is In field.
No Seasonality	Indicates that your data are not seasonal. If you select this setting, CB Predictor doesn't use the seasonal methods in the Method Gallery.

Use Multiple Linear Regression	<p>Specifies that you have at least one data series that depends on one or more other data series. Selecting this setting automatically opens the Regression Variables dialog. For more information, see “Regression Variables dialog” on page 106.</p> <p>For example, if you want to predict national sales of big-screen televisions, you might know that it depends on National Personal Income and Savings and Loans Deposits. Therefore, you should use multiple linear regression as the forecasting method for the sales of big-screen televisions.</p>
Select Variables	<p>Opens the Regression Variables dialog, which lets you identify your dependent and independent variables.</p>
Method	<p>Sets what regression method to use. The choices are standard, forward stepwise, and iterative stepwise.</p> <p>For more information on these regression methods, see “Regression methods” on page 48.</p> <p>The default is Standard.</p>
Stepwise Options	<p>Opens the Stepwise Options dialog. If you select a stepwise regression method, you can use this button to set the stopping criteria for the stepwise regression or the partial F statistic significance to use for the steps.</p> <p>For more information, see “Stepwise Options dialog” on page 108.</p>
Include Constant	<p>Sets whether CB Predictor must generate a regression equation with or without a constant. No constant indicates that the y-intercept of the regression line is 0.</p> <p>The default is to use a constant. Turn off this setting only if the dependent variable is zero when all independent variables are zero.</p>

Regression Variables dialog

You can use the Regression Variables dialog to select the dependent and independent variables for forecasting with multiple linear regression and to set settings for independent variable series. To display this dialog, click Select Variables in Step 5 of the Data Attributes tab (Figure A.4, page 105).

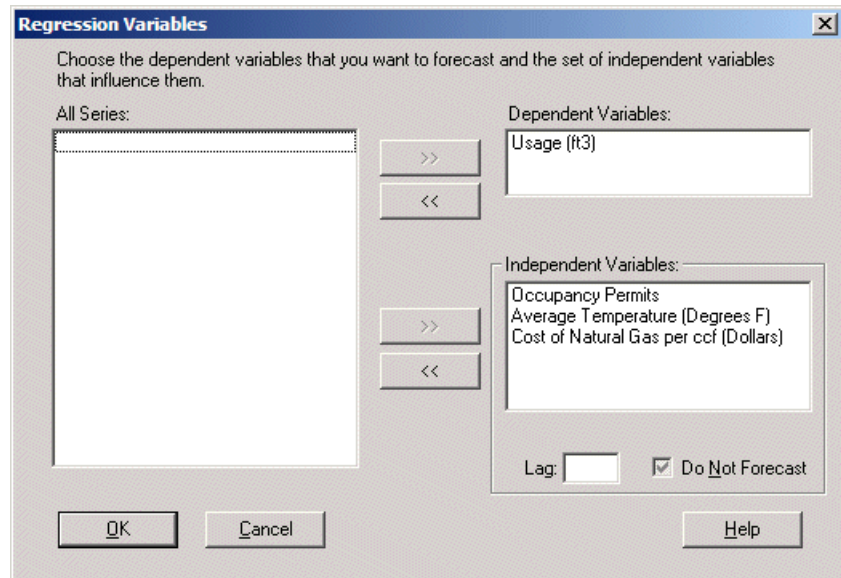


Figure A.5 Regression Variables dialog

The fields, settings, and buttons in this dialog are:

All Series	Lists all the variables on your spreadsheet that are not designated as either dependent or independent variables.
Dependent Variables	Lists all the variables selected as being dependent on other variables. You must select at least one dependent variable.
Independent Variables	Lists all the variables selected as having influence on dependent variables. You must select at least one independent variable if you selected a dependent variable.

Note: *You don't have to define all your variables as either dependent or independent. Just leave variables that are not involved in the regression in the All Series list.*

Lag	Sets the number of periods that you want the dependent variable to follow after the independent variable for calculating the regression equation. The default is 0.
-----	--

Do Not Forecast	<p>Sets whether to forecast the independent variable during Hypercasting. You might not want to forecast some independent variables if you have future values determined in some external way and already have those values in your spreadsheet before forecasting.</p> <p>The default is off, meaning that CB Predictor forecasts all independent variables during Hypercasting.</p>
>>	Moves the variables selected in the All Series list to either the corresponding Dependent Variables or Independent Variables list.
<<	Moves the variables selected in either the Dependent Variables or Independent Variables list to the All Series list.

Stepwise Options dialog

You can set settings that control how CB Predictor does stepwise regression in the Stepwise Options dialog.

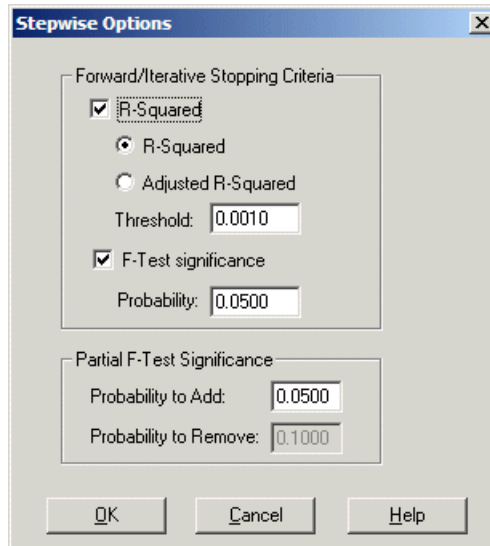


Figure A.6 Stepwise Options dialog

The fields, settings, and buttons in the Stepwise Options dialog are:

R -Squared	<p>Stops the stepwise regression if the difference between a specified statistic (either R^2 or adjusted R^2) for the previous and new regression solutions is below a threshold value. When this happens, CB Predictor does not use the new regression solution.</p> <p>By default, this stopping criterion is on and uses R^2 as the statistic. If this setting and the F statistic significance are both on, the stepwise regression stops when it reaches either criterion's threshold value.</p>
Threshold	<p>Sets the minimum increment required between the R^2 (or adjusted R^2) of the last step and the R^2 (or adjusted R^2) of the new step to continue with the stepwise regression.</p> <p>The default is 0.001.</p>
F -test significance	<p>Stops the stepwise regression if the probability of the F statistic for a new solution is above a maximum value.</p> <p>By default, this stopping criterion is on. If this setting and the R-squared setting are both on, the stepwise regression stops when it reaches either criterion's threshold value.</p>
Probability	<p>Sets the maximum probability required for the F statistic to continue with the stepwise regression.</p> <p>The default is 0.05.</p>
Probability To Add	<p>Sets the maximum probability of the correlation (partial F statistic) of the independent variable required to add the variable to the regression equation.</p> <p>The default is 0.05.</p>
<hr/>	
	<p>Note: When dealing with statistical tests, smaller probabilities indicate more significance.</p>
Probability To Remove	<p>Sets the minimum probability of the correlation (partial F statistic) of the independent variable required to remove the variable from the regression equation.</p> <p>The default is 0.1.</p>

Note: This setting is only available with iterative stepwise regression. The Probability To Remove must be at least 0.05 higher than the Probability To Add.

Method Gallery tab

The Method Gallery tab lets you select any combination of the time-series forecasting methods to try when forecasting your data. Selecting extra methods doesn't significantly increase your forecast time unless you have thousands of data values.

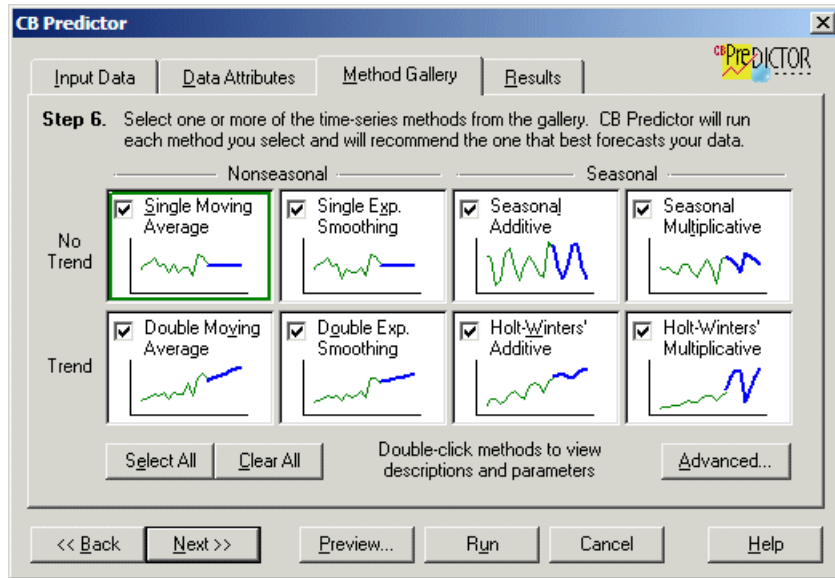


Figure A.7 Method Gallery tab

Each method area has a:

Title	The title of the method. For more information on each method, see “Selecting time-series forecasting methods” on page 66 or Chapter 2, “Understanding the Terminology.”
Check box	Lets you select to try that method. You can select anywhere between one and all of the methods.
Sample graph	Shows what sample data and fitted data might look like for each method.

In addition to this, double-clicking on each time-series method area opens that method's parameters dialog. The parameters dialog lets you view a description and a sample graph of each method and manually set the parameters for that method. You don't need to set the parameters for the

time-series forecasting methods. CB Predictor automatically finds the optimal parameters unless you override this optimization by selecting to use User Defined parameters in this dialog. For more information on this dialog, see “Parameters dialogs” on page 111.

The other buttons and settings for this tab are:

Select All	Selects all the time-series forecasting methods.
Clear All	Clears all the method checkboxes.
Advanced	Opens the Advanced dialog, which lets you select the error measure to determine which method best fits the historical data. For more information, see “Advanced Options dialog” on page 113.

Parameters dialogs

The parameters dialogs all have the same format. To display this dialog, double-click a time-series method in the Method Gallery tab (Figure A.7, page 110).

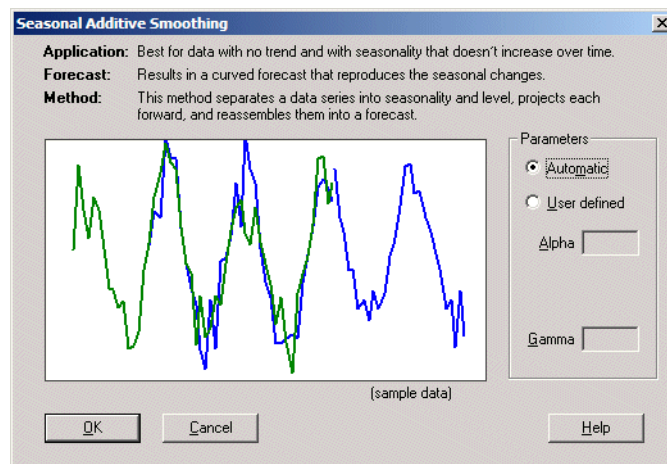


Figure A.8 Parameters dialog for the Seasonal Additive Smoothing method

Parameters dialogs all include the following fields and settings:

Title	The full title of method. Some titles on the Method Gallery dialog are abbreviated.
Application	Describes for what kind of data the method works best.
Forecast	Describes what the resulting forecast looks like.
Method	Describes how the method calculates its forecast.
Chart	Shows an example of fitted data with sample data.
Automatic	Sets CB Predictor to automatically optimize the values for the method's parameters. This is the default.
User Defined	<p>Lets you enter values in the parameter fields. See Appendix B, "Time-Series Forecasting Method Formulas," for more information on these parameters.</p> <p>CB Predictor will use these user-defined values on all forecasts until you select Automatic again. A double asterisk appears next to the name in the Method Gallery for all methods using user-defined parameters.</p>
Periods	<p>Sets the number of periods that the method uses. You can type in this field only if you select User Defined.</p> <p>The default is 3 for single moving average and 6 for double moving average. This parameter is available only for these methods.</p>
Alpha	<p>Sets the Alpha parameter for the method, if applicable. You can type in this field only if you select User Defined.</p> <p>This parameter is <i>not</i> available for the single moving average or double moving average methods.</p>
Beta	<p>Sets the Beta parameter for the method, if applicable. You can type in this field only if you select User Defined.</p> <p>This parameter is available only for the double exponential smoothing, Holt-Winters' additive and Holt-Winters' multiplicative methods.</p>
Gamma	<p>Sets the Gamma parameter for the method, if applicable. You can type in this field only if you select User Defined.</p> <p>This parameter is available only for the seasonal methods.</p>

Advanced Options dialog

The Advanced Options dialog lets you select which error measure to use to determine which forecasting method best fits your data, and which forecasting technique to use to perform time-series forecasting. To display this dialog, click Advanced near the bottom of the Method Gallery tab (Figure A.7, page 110).

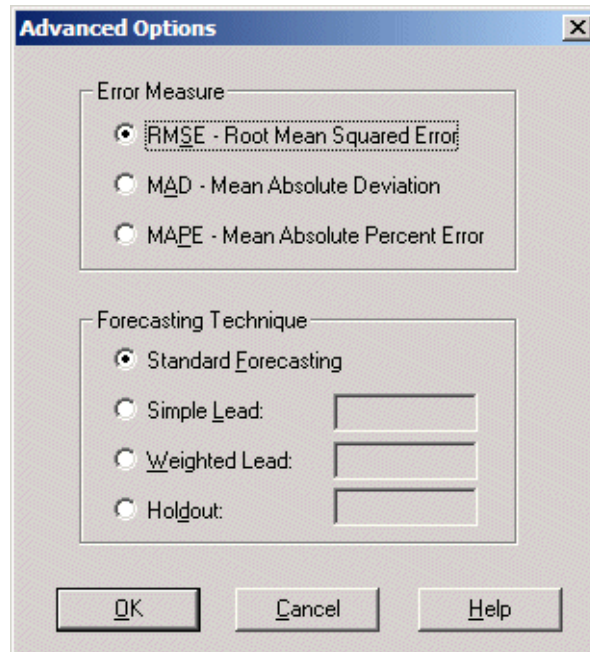


Figure A.9 Advanced Options dialog

The settings on this dialog are:

RMSE	Selects root mean squared error as the statistic that determines the best forecast. This is the default error measure. For more information, see “RMSE” on page 42 and Appendix C.
MAD	Selects mean absolute deviation as the statistic that determines the best forecast. For more information, see “MAD” on page 43 and Appendix C.

MAPE	<p>Selects mean absolute percentage error as the statistic that determines the best forecast.</p> <p>For more information, see “MAPE” on page 43 and Appendix C.</p>
Standard Forecasting	<p>Selects to use the standard forecasting technique for time-series forecasting. This is the default forecasting technique.</p> <p>For more information, see “Time-series forecasting techniques” on page 43.</p>
Simple Lead	<p>Selects to use simple lead forecasting for performing time-series forecasting. If you select this technique, enter the number of lead periods you want to use in the adjacent field.</p> <p>The default number of lead periods is 1.</p> <p>For more information, see “Simple lead forecasting” on page 44.</p>
Weighted Lead	<p>Selects to use weighted lead forecasting for performing time-series forecasting. If you select this technique, enter the number of weighted lead periods you want to use in the adjacent field.</p> <p>The default number of weighted lead periods is 1.</p> <p>For more information, see “Weighted lead forecasting” on page 44.</p>
Holdout	<p>Selects to use the holdout technique for time-series forecasting. If you select this technique, enter the number of periods you want to exclude from the forecast.</p> <p>The default number of periods to hold out is 1.</p> <p>For more information, see “Holdout” on page 45.</p>

Results tab

The Results tab lets you select what kind of output you want CB Predictor to generate for you.

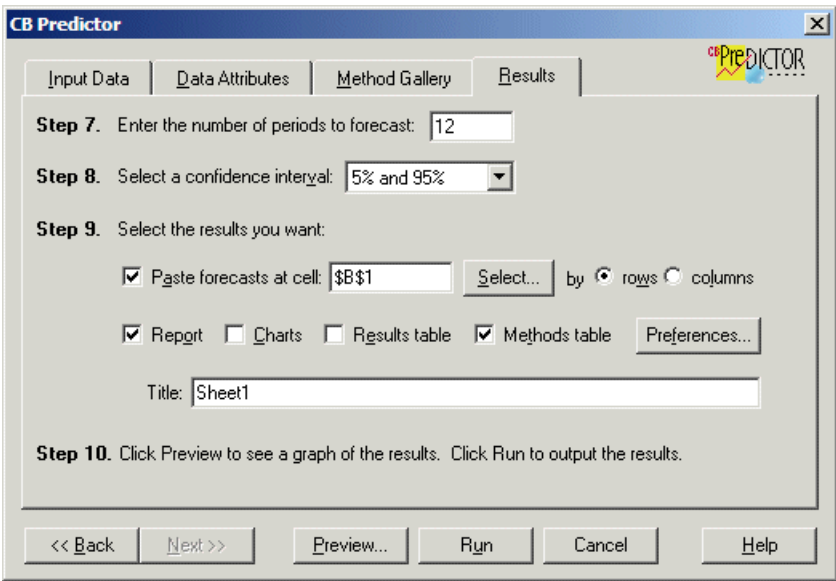


Figure A.10 Results tab of the wizard dialog

The settings, fields, and buttons of the Results tab are:

Number Of Periods To Forecast	Lets you enter the number of time periods you want CB Predictor to forecast for each series. You can enter any integer from 1 to 100, but don't use a value larger than the number of historical data points. The default is 4.
Confidence Interval	Lets you select the probabilities of the confidence interval from a list of typical ones. The default is 5% and 95%.

Paste Forecasts	<p>Selects to add the forecasted values to the end of the historical data range on your original spreadsheet or somewhere else in your workbook.</p> <p>By default, this is on, set to paste points at the end of your historical data. To select a cell range different from the one in the Paste field, click Select to change the location and select to paste the data either in rows or in columns.</p>
Select	<p>Brings up a small Select Range dialog to let you select the cells directly from the spreadsheet. After selecting the cells, click OK to return to the Results tab.</p>
Rows/Columns	<p>Defines whether you want your data pasted in rows or columns.</p>
Report	<p>Creates a customizable, detailed report that can include elements such as a summary for each forecasted variable, which method was best, what parameters the method used, and the error.</p> <p>This is off by default.</p>
Chart	<p>Creates a chart for each data series to forecast. Each chart shows, optionally, the historical data, fitted values, forecasted values, and a confidence interval.</p> <p>You can put only 40 CB Predictor charts on a workbook. If you need more than 40 charts, CB Predictor creates new workbooks when necessary.</p> <p>This is off by default.</p>
Results Table	<p>Creates a customizable PivotTable that lists items such as all the forecasted and fitted values, as well as the values defining the confidence interval.</p> <p>This is off by default.</p>
Methods Table	<p>Creates a customizable PivotTable that lists items such as each method tried for each variable, as well as all the errors, parameters, and statistics for each method.</p> <p>This is off by default.</p>
Preferences	<p>Lets you set the different results' settings, such as whether to prompt you when pasting will overwrite data or whether to include charts on your reports. This button opens the Preferences dialog, with tabs for each different result.</p> <p>For more information, see "Preferences dialog" on page 117.</p>
Title	<p>Lets you enter a title that will appear for most results.</p> <p>The default is the worksheet name.</p>

Preferences dialog

The Preferences dialog lets you set settings about how your selected results appear. To display this dialog, click Preferences in Step 9 of the Results tab (Figure A.10, page 115).

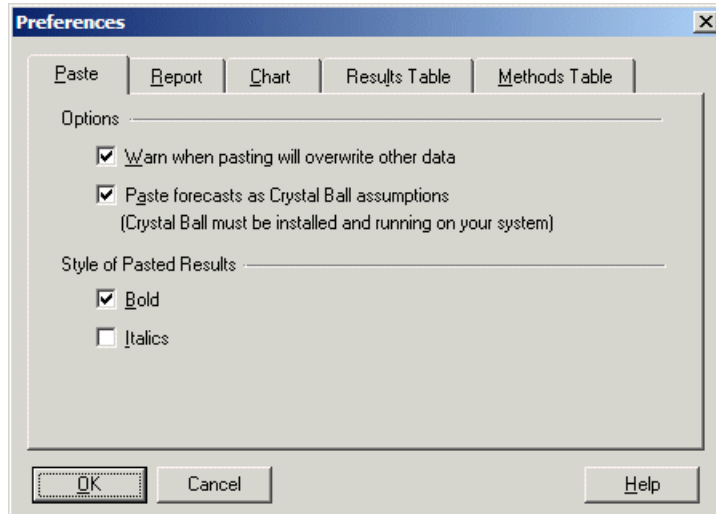


Figure A.11 Preferences dialog

There are tabs for each type of result:

- Paste (page 118)
- Report (page 119)
- Chart (page 121)
- Results Table (page 122)
- Methods Table (page 123)

CB Predictor Note: You can change results settings for unselected result settings. You still don't generate a type of result unless you select it in the Results tab of the wizard dialog.

Preferences dialog, Paste tab

The settings for the Paste Forecasts result are on the Preferences dialog under the Paste tab.

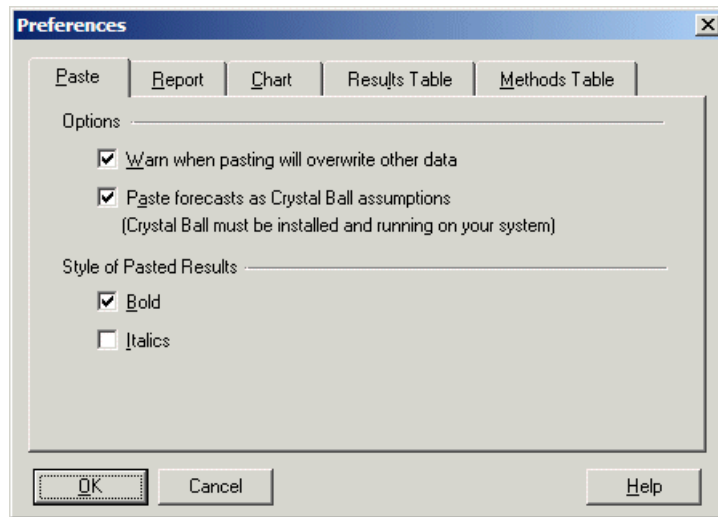


Figure A.12 Preferences dialog — Paste tab

You can change these settings even if Paste Forecasts isn't a selected result. CB Predictor saves the settings and applies them the next time you use the Paste Forecasts result.

The settings in this tab are:

Warn When Pasting Will Overwrite Other Data	Displays a warning when your pasted data will overwrite existing data on your spreadsheet. This is on by default.
Paste Forecasts As Crystal Ball Assumptions	Creates pasted cells as Crystal Ball assumptions. The defined assumptions are normal distributions with a mean that is the forecasted value and a standard deviation that is based on the RMSE of the fitted data.

Note: CB Predictor doesn't create assumptions if the variation in the data is zero or approaches infinity.

This is on by default.

Style Of Pasted Results	Sets how the pasted data appear on your spreadsheet. This helps you identify the forecasted data from your historical data. You can set the forecasted data to be bold, italics, or both. The default is bold only.
-------------------------	---

Preferences dialog, Report tab

The settings for the Report result are on the Preferences dialog under the Report tab.

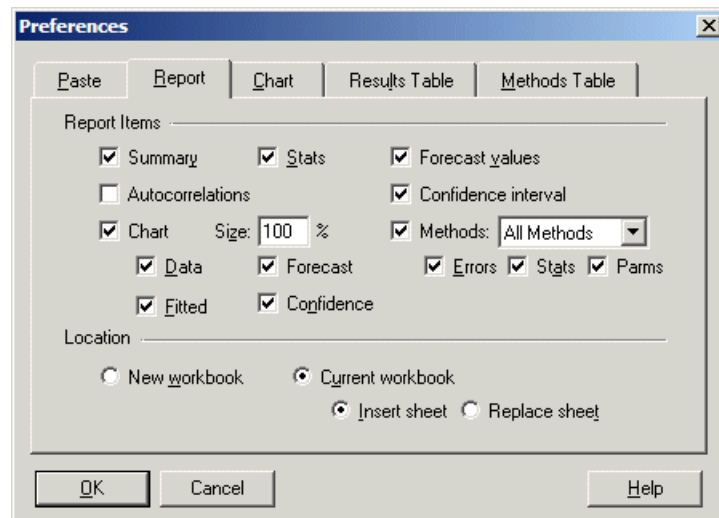


Figure A.13 Preferences dialog — Report tab

You can change these settings even if Report isn't a selected result. CB Predictor saves the settings and applies them the next time you use the Report result.

The settings in this tab are:

Summary	Includes a summary section at the top of the report. The report includes a summary by default.
Autocorrelations	Includes the autocorrelation coefficients for the top 18 autocorrelations for each data series. The report does not include autocorrelation by default.

Chart	<p>Includes charts in the report. You can also set your chart settings differently than your regular chart settings. For more information on the chart settings, see “Preferences dialog, Chart tab” on page 121.</p> <p>The report includes charts by default.</p>
Stats	<p>Includes the series statistics (e.g., mean, standard deviation) and, if the series is a dependent variable, the regression statistics (e.g., R^2, Durbin-Watson).</p> <p>The report includes statistics by default.</p>
Forecast Values	<p>Includes the actual forecasted values.</p> <p>The report includes forecasted values by default.</p>
Confidence Interval	<p>Includes the points that define the confidence interval.</p> <p>The report includes confidence interval points by default. If you selected None for the confidence interval on the Results tab, no confidence interval appears in the report.</p>
Methods	<p>Displays information about the methods CB Predictor tried for each series. You can display information for all the methods tried or for only the top one, two, or three methods. For each method, you can select to display the errors, statistics, parameters, or any combination.</p> <p>By default, the report includes all the information about all the methods.</p>
Location	<p>Indicates whether the report should appear in a new worksheet of the current workbook or in a new workbook.</p> <p>If you select to place reports in the current workbook, you can also indicate whether you want CB Predictor to create a new worksheet every time you run your forecast or to replace the report on the worksheet when you rerun your forecast.</p> <p>The default is Current Workbook with Insert Sheet.</p>

Preferences dialog, Chart tab

The settings for the Chart result are on the Preferences dialog under the Chart tab.

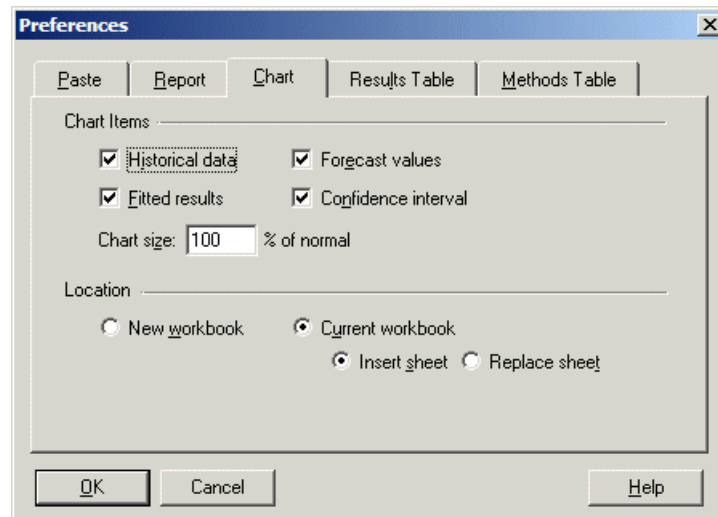


Figure A.14 Preferences dialog — Chart tab

You can change these settings even if Chart isn't a selected result. CB Predictor saves the settings and applies them the next time you use the Chart result.

The settings in this tab are:

- | | |
|-------------|--|
| Chart Items | Selects which types of data appear on the charts. You can select from any combination of Historical Data, Fitted Results, Forecast Values, and Confidence Interval.
By default, all the items appear on the charts. |
| Chart Size | Scales the generated charts to a percentage of the normal size. You can enter the percentage as an integer between 20 and 400. The default is 100. |

Location	<p>Indicates whether the charts should appear in a new worksheet of the current workbook or in a new workbook.</p> <p>If you select to place the charts in the current workbook, you can also indicate whether you want CB Predictor to create a new worksheet every time you run your forecast or to replace the charts on the worksheet when you rerun your forecast.</p> <p>The default is Current Workbook with Insert Sheet.</p>
----------	---

Preferences dialog, Results Table tab

The settings for the Results Table result are on the Preferences dialog under the Results Table tab.

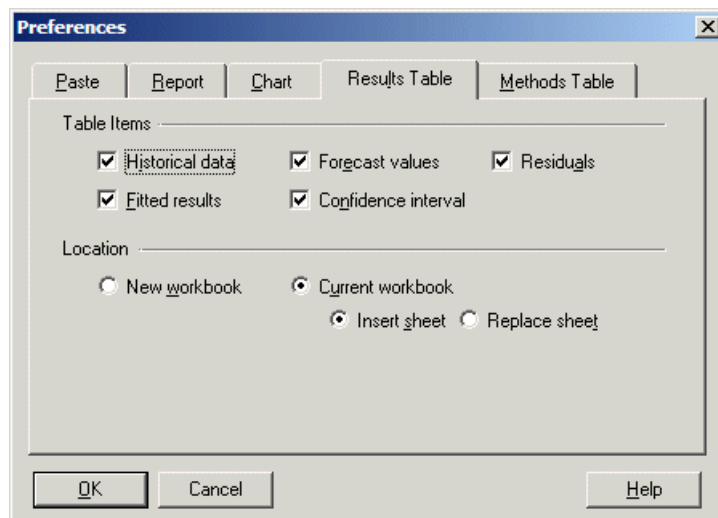


Figure A.15 Preferences dialog — Results Table tab

You can change these settings even if Results Table isn't a selected result. CB Predictor saves the settings and applies them the next time you use the Results Table result.

The settings in the Results Table tab are:

Table Items	Selects which items to include in the Results table. You can include any combination of Historical Data, Fitted Results, Forecast Values, Confidence Interval, and Residuals. The default is to include all the items.
Location	Indicates whether the PivotTable should appear in a new worksheet of the current workbook or in a new workbook. If you select to place the results table in the current workbook, you can also indicate whether you want CB Predictor to create a new worksheet every time you run your forecast or to replace the results table on the worksheet when you rerun your forecast. The default is Current Workbook with Insert Sheet.

Preferences dialog, Methods Table tab

The settings for the Methods Table result are on the Preferences dialog under the Methods Table tab.

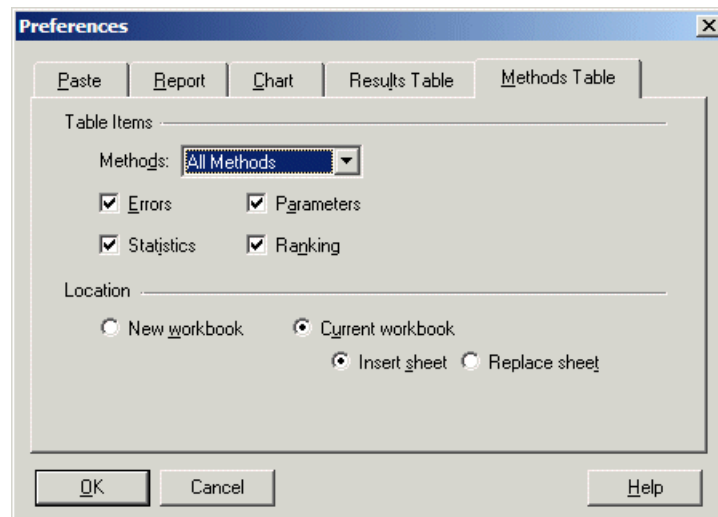


Figure A.16 Preferences dialog — Methods Table tab

You can change these settings even if Methods Table isn't a selected result. CB Predictor saves the settings and applies them the next time you use the Methods Table result.

The settings in the Methods Table tab are:

Table Items	<p>Selects which items to include in the Methods table. You can include information on all the forecasting methods that CB Predictor tried or information on only the top one, two, or three methods.</p> <p>For each method, you can include any combination of Errors, Statistics, Parameters, and Ranking.</p> <p>The default is to include all the information for all the forecasting methods.</p>
Location	<p>Indicates whether the PivotTable should appear in a new worksheet of the current workbook or in a new workbook.</p> <p>If you select to place the generated methods table in the current workbook, you can also indicate whether you want CB Predictor to create a new worksheet every time you run your forecast or to replace the methods table on the worksheet when you rerun your forecast.</p> <p>The default is Current Workbook with Insert Sheet.</p>

Preview Forecast dialog

You can open the Preview dialog from the wizard at any time after you define valid input data. To display this dialog, click Preview at the bottom of the Results tab (Figure A.10, page 115).

To generate values for the Preview dialog, CB Predictor actually calculates the forecast for each series using each of the methods selected in the Method Gallery tab of the wizard. These results appear in the Preview chart.

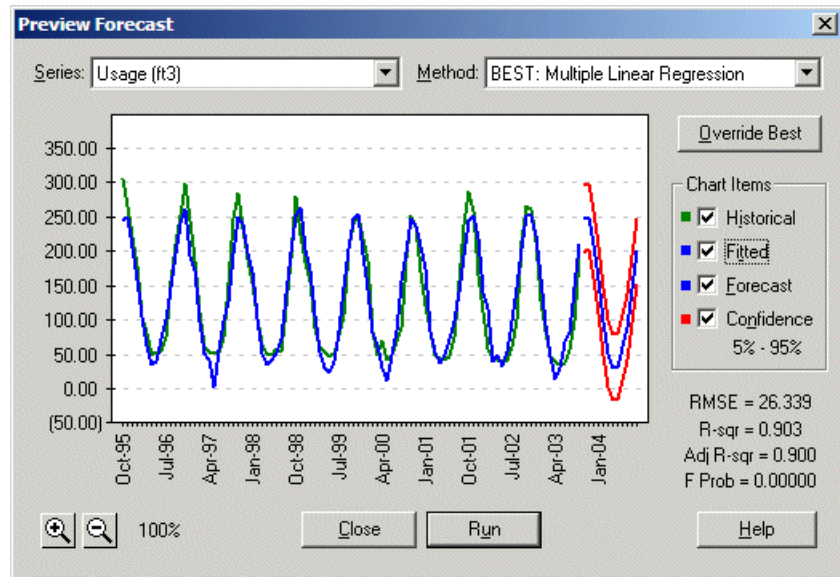




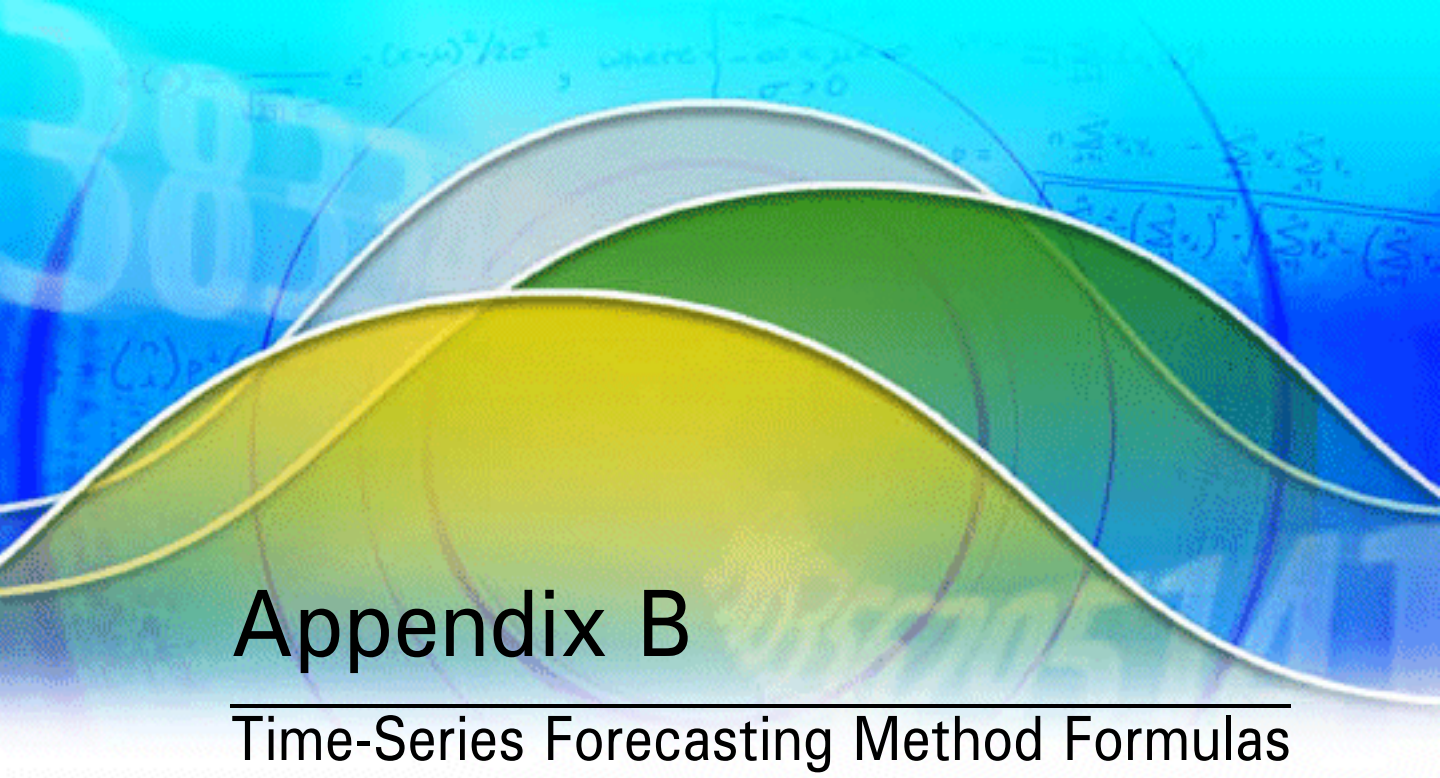
Figure A.17 Preview dialog

The Preview Forecast chart displays the historical, fitted, forecast, and confidence interval data values for each method and for each data series. To the right of the chart, it also displays the error measure and the parameters (as applicable) for the selected series and method.

In addition to previewing your data, the preview dialog lets you override the method that CB Predictor found to forecast best. When you click Run in the forecast wizard, CB Predictor outputs results to a new or current worksheet. The results are for the method that best fits your historical data. If, after viewing the forecast from the different methods, you decide to use the results of another method, you can override this best method. CB Predictor then outputs the results using this new best method.

The fields, settings, and buttons in the Preview dialog are:

Series	Lists all the data series in your spreadsheet. The selected series appears in the chart.
Method	<p>Lists from best to worst all the data methods that CB Predictor tried when forecasting the selected series from the Series field. The selected method is used to forecast the selected series (from the Series field) in the chart.</p> <p>If you indicated on the Data Attributes tab that your data had no seasonality, CB Predictor skips the seasonal methods, and they do not appear in this list, even if you selected them in the Method Gallery.</p>
Override Best	<p>Lets you reassign a new method to be the “best.” The new best method appears in the list as the best. The best method (that the program chose or that you overrode) is used for all output results.</p> <p>Use this button to override the best method if your visual inspection tells you that another method is better. For example, sometimes CB Predictor selects a moving average method for seasonal data because it fits the historical data closely, even though it wouldn’t predict future values the best.</p>
Chart	Displays all the values graphically for the items selected in the Chart Items area.
Chart Items	Selects the items to display in the chart. The choices are Historical, Fitted, Forecast, and Confidence.
Errors and parameters	Displays the error (using the error measure in the Advanced dialog) for the current series and method. If the method has associated parameters, they appear below the error.
 Zoom In	Increases the magnification of the data graph, showing fewer data points.
 Zoom Out	Decreases the magnification of the data graph, showing more data points.



Appendix B

Time-Series Forecasting Method Formulas

In this appendix

This appendix provides formulas for the following time-series forecasting methods:

- Double moving average
- Single exponential smoothing
- Holt's double exponential smoothing
- Seasonal additive smoothing
- Seasonal multiplicative smoothing
- Holt-Winters' additive seasonal smoothing
- Holt-Winters' multiplicative seasonal smoothing

Nonseasonal forecasting method formulas

Below are the formulas of the nonseasonal time-series forecasting methods.

CB Predictor Note: This appendix doesn't list the formula for single moving average because of its simple nature.

Double moving average

The formula CB Predictor uses for the double moving average method is:

$$2 * S[1](t) - S[2](t) + \frac{2}{p-1} (S[1](t) - S[2](t))$$

where $S[1](t)$ represents the moving average (with period p) of the historical data, $S[2](t)$ represents the moving average of the $S[1](t)$ series (with period p), and p represents the number of time periods the moving averages are over.

Single exponential smoothing

The formula CB Predictor uses for the single exponential smoothing method is:

$$S[1](t) = \alpha * y(t-1) + (1 - \alpha) * S[1](t-1)$$

where $S[1](t)$ represents the single exponential smoothing “smoothed estimate” for time period t , $y(t)$ is the historical data value at time t , and α is the smoothing constant between 0 and 1.

Holt's double exponential smoothing

CB Predictor uses Holt's double exponential smoothing instead of Brown's double exponential smoothing. The difference is that when the single exponential smoothing equation is applied the second time, Holt's method lets you use a different smoothing constant (β), compared to Brown's method which uses the same constant for both operations.

The formula for Holt's double exponential smoothing is:

$$S[2](t) = \beta * (S[1](t) - S[1](t-1)) + (1 - \beta) * S[2](t-1)$$

where $S[2](t)$ represents the double exponential smoothing “smoothed estimate” for time period t , $S[1](t)$ represents the single exponential smoothing “smoothed estimate” for time period t , $y(t)$ is the historical data

value at time t , and α and β are the smoothing constants with values between 0 and 1.

Seasonal forecasting method formulas

Below are the formulas for the seasonal time-series forecasting methods.

Seasonal additive smoothing

CB Predictor uses the following equations to calculate this method:

$$(\text{Level}) L_t = \alpha * (Y_t - S_{t-s}) + (1 - \alpha) * L_{t-1}$$

$$(\text{Seasonal}) S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-s}$$

$$(\text{Forecast for period } m) F_{t+m} = L_t + S_{t+m-s}$$

where the parameters are:

α	alpha
γ	gamma
m	the number of periods ahead to forecast
s	the length of the seasonality
L_t	the level of the series at time t
S_t	the seasonal component at time t

Seasonal multiplicative smoothing

CB Predictor uses the following equations to calculate this method:

$$(\text{Level}) L_t = \alpha * (Y_t / S_{t-s}) + (1 - \alpha) * L_{t-1}$$

$$(\text{Seasonal}) S_t = \gamma * (Y_t / L_t) + (1 - \gamma) * S_{t-s}$$

$$(\text{Forecast for period } m) F_{t+m} = L_t * S_{t+m-s}$$

where the parameters are:

α	alpha
γ	gamma
m	the number of periods ahead to forecast

s	the length of the seasonality
L_t	the level of the series at time t
S_t	the seasonal component at time t

Holt-Winters' additive seasonal smoothing

$$(\text{Level}) L_t = \alpha * (Y_t - S_{t-s}) + (1 - \alpha) * (L_{t-1} + b_{t-1})$$

$$(\text{Trend}) b_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1}$$

$$(\text{Seasonal}) S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-s}$$

$$(\text{Forecast for period } m) F_{t+m} = L_t + m * b_t + S_{t+m-s}$$

where the parameters are:

α	alpha
β	beta
γ	gamma
m	the number of periods ahead to forecast
s	the length of the seasonality
L_t	the level of the series at time t
b_t	the trend of the series at time t
S_t	the seasonal component at time t

Holt-Winters' multiplicative seasonal smoothing

$$\text{(Level) } L_t = \alpha * (Y_t / S_{t-s}) + (1 - \alpha) * (L_{t-1} + b_{t-1})$$

$$\text{(Trend) } b_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1}$$

$$\text{(Seasonal) } S_t = \gamma * (Y_t / L_t) + (1 - \gamma) * S_{t-s}$$

$$\text{(Forecast for period m) } F_{t+m} = (L_t + m * b_t) * S_{t+m-s}$$

where the parameters are:

α	alpha
β	beta
γ	gamma
m	the number of periods ahead to forecast
s	the length of the seasonality
L_t	the level of the series at time t
b_t	the trend of the series at time t
S_t	the seasonal component at time t



Appendix C

Error Measure and Statistic Formulas

In this appendix

This appendix provides formulas for the following types of statistics used in CB Predictor:

- Time-series forecast error measures
- Confidence intervals
- Time-series forecast statistics
- Autocorrelation statistics

Time-series forecast error measures

CB Predictor calculates three different error measures for the fit of each time-series forecast. CB Predictor uses one of these error measures to determine which time-series forecasting method is the best.

RMSE

Root mean squared error. This is an absolute error measure that squares the deviations to keep the positive and negative deviations from cancelling out each other. This measure also tends to exaggerate large errors, which can help when comparing methods.

The formula for calculating RMSE is:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n}}$$

where Y_t is the actual value of a point for a given time period t , n is the total number of fitted points, and \hat{Y}_t is the fitted forecast value for the time period t . For example, the RMSE for the sample data plotted in Figure 2.9 is 0.3068.

To calculate the RMSE for a set of data:

1. **Measure the difference between each forecasted point (\hat{Y}_t) and its corresponding actual point (Y_t).**
2. **Square the difference.**
3. **Repeat steps 1 through 2 for each fitted point and sum the results.**
4. **Divide the sum by the number of measured differences.**
5. **Take the square root.**

MAD

Mean absolute deviation. This is an error statistic that averages distance between each pair of actual and fitted data points.

The formula for calculating the MAD is:

$$\text{MAD} = \frac{\sum_{t=1}^n |Y_t - \hat{Y}_t|}{n}$$

where Y_t is the actual value of a point for a given time period t , n is the total number of fitted points, and \hat{Y}_t is the forecast value for the time period t . For example, the MAD for the sample data plotted in Figure 2.9 is 0.2799.

To calculate the MAD for a set of data:

1. **Measure the difference between each forecasted point (\hat{Y}_t) and its corresponding actual point (Y_t).**
2. **Take the absolute values of these differences.**
3. **Add up all the absolute values of the differences.**
4. **Divide the sum of the differences by the number of measured deviations (n).**

MAPE

Mean absolute percentage error. This is a relative error measure that uses absolute values to keep the positive and negative errors from cancelling out each other and uses relative errors to let you compare forecast accuracy between time-series models.

The formula for calculating the MAPE is:

$$\text{MAPE} = \frac{\sum_{t=1}^n \left| \frac{(Y_t - \hat{Y}_t)}{Y_t} (100) \right|}{n}$$

where Y_t is the actual value of a point for a given time period t , n is the total number of fitted points, and \hat{Y}_t is the forecast value for the time period t .

CB Predictor Note: If Y_t equals zero, CB Predictor drops the term $\frac{(Y_t - \hat{Y}_t)}{Y_t}$.

For example, the MAPE for the sample data plotted in Figure 2.9 is 138.17%.

To calculate the MAPE for a set of data:

1. **Measure the difference between each forecasted point (\hat{Y}_t) and its corresponding actual point (Y_t).**
2. **Divide the difference by the actual value.**
3. **Multiply the result of step 2 by 100.**
4. **Take the absolute value of this product.**
5. **Repeat steps 1 through 4 for each fitted point and sum the results.**
6. **Divide the sum by the number of measured differences.**

Confidence intervals

The confidence interval defines the range within which a forecasted value has some probability of occurring. CB Predictor used the empirical method of calculating confidence intervals.

If we define y_1 as the first of a set of historical data with a total of n observations, suppose that at time period t we start forecasting data m periods into the future. Then, we denote the m -period-ahead forecast error at time period t as $e_t(m)$.

The standard error of prediction (S) is equal to the square root of the mean of the squared forecast errors, or

$$S(m) = \sqrt{\frac{\sum_{t=1}^{n-m} e_t^2(m)}{n-m}}$$

Assuming that forecast errors are normally distributed, the formula for predicting the future value of y_{t+m} at time t within a 95% confidence interval is

$$\hat{y}_{t+m}(t) \pm 1.959996S(m)$$

The empirical method is reasonably accurate as long as the amount of historical data is sufficiently large. For further information, see page 427 ff. of Bowerman, B.L., and R.T. O'Connell, 1993 (see “Bibliography” on page 159).

Time-series forecast statistics

Another statistic calculated for any time-series forecast is Theil's U .

Theil's U statistic

This statistic, described on page 46, compares forecasted results with the results of forecasting with minimal historical data.

The formula for calculating Theil's U statistic is:

$$U = \sqrt{\frac{\sum_{t=1}^{n-1} \left(\frac{\hat{Y}_{t+1} - Y_{t+1}}{Y_t} \right)^2}{\sum_{t=1}^{n-1} \left(\frac{Y_{t+1} - Y_t}{Y_t} \right)^2}}$$

where Y_t is the actual value of a point for a given time period t , n is the number of data points, and \hat{Y}_t is the forecasted value.

To calculate the U statistic for a set of data:

1. **Starting with the second time period, subtract the actual from the forecasted \hat{Y}_t .**
2. **Divide the result by the actual Y of the previous time period.**
3. **Square the result.**
4. **Repeat steps 1 to 3 for the rest of the data, except for the last one, and add the results.**
5. **Starting with the second time period, subtract the actual Y of the previous time period from the forecasted Y of the current time period.**
6. **Divide the result by the actual Y of the previous time period.**
7. **Square the result.**
8. **Repeat steps 5 to 7 for the rest of the data, except for the last one, and add the results.**

9. Divide the result of step 4 by the result of step 8.
10. Take the square root of the result in step 9.

Autocorrelation statistics

Measures of autocorrelation describe the relationship among values of the same data series at different time periods.

Durbin-Watson

The Durbin-Watson statistic, described on page 51, calculates autocorrelation at lag 1.

The formula for calculating the Durbin-Watson statistic is:

$$DW = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2}$$

where e_t is the difference between the estimated point (\hat{Y}_i) and the actual point (Y_i) and n is the number of data points.

To calculate the Durbin-Watson statistic for a set of data:

1. For each point, calculate the error by subtracting the estimated point (\hat{Y}_i) from the actual point (Y_i).
2. Starting with the second data point, subtract from the error the error of the previous data point.
3. Square the difference.
4. Repeat steps 2 and 3 for the rest of the data points.
5. Add the squares.
6. Starting with the first data point, square each error.
7. Add up the squares.
8. Divide the result of step 5 by the result of step 7.

Autocorrelation

Measuring autocorrelation can help detect seasonality in historical data, as described in “View Historical Data dialog” on page 102. CB Predictor uses the following formula to estimate autocorrelation:

$$r_k = \frac{\sum_{t=k+1}^n (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=k+1}^n (y_t - \bar{y})^2}$$

where y is the data series, y_t is the actual value of a point at time t , n is the number of data points, k is the lag between correlated values of the series, and \bar{y} is the sample mean of the series.

Ljung-Box statistic

This statistic measures whether a set of autocorrelations is significantly different from a set of autocorrelations that are all zero. The Ljung-Box statistic is calculated as follows:

$$Q' = n(n+2) \sum_{k=1}^{h-1} \frac{r_k^2}{(n-k)}$$

where:

Q'	Is the Ljung-Box statistic, the probability that the set of autocorrelations is the same as a set of autocorrelations which are all 0.
n	Is the amount of data in the data sample.
h	Is the size of the set of autocorrelations used to calculate the statistic.
rk	Is the autocorrelation with a lag of k .

The size of the set of autocorrelations is equal to 1/3 the size of the data sample or 100, if the sample is larger than 300.



Appendix D

Techniques for Finding Regression Coefficients

In this appendix

- Least squares
- Singular value decomposition

This appendix describes the two most common methods for finding the coefficients of a linear regression equation: least squares and singular value decomposition.

It also describes why CB Predictor uses singular value decomposition instead of the least squares approach to determine the multiple linear regression equations.

Finding regression coefficients

There are two common techniques used to find the coefficients of a regression equation that best fit the historical data:

- Least squares
- Singular value decomposition

Of these two techniques, CB Predictor uses singular value decomposition. The following sections describe both techniques and describe why CB Predictor uses the singular value decomposition technique.

Least squares

Least squares is a technique that estimates the coefficients of an equation that minimizes the SSE. There are several ways to calculate the least squares coefficients, such as using partial differential equations or matrices.

For the matrix technique, use the equation:

$$y = b_0 + b_1x_{t1} + b_2x_{t2} + b_3x_{t3} + \dots + b_px_{tp} + e$$

Which can be written in the form:

$$\mathbf{y} = \mathbf{X} * \mathbf{b}$$

where for n time periods and p independent variables, the matrices \mathbf{y} and \mathbf{X} are defined as:

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \text{and} \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

Solving this equation using matrix algebra, gives the coefficients \mathbf{b} as:

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

where \mathbf{X}' is the transpose of the matrix \mathbf{X} and $(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of the matrix $\mathbf{X}'\mathbf{X}$.

Singular value decomposition

Singular value decomposition is also technique for determining the coefficients of a regression equation. The primary difference between the singular value decomposition and the least squares techniques is that the singular value decomposition technique can handle situations where the equations used to determine the coefficients of the regression equation are singular or close to singular. This happens when performing regression on equations that represent parallel lines or surfaces.

In these cases, the least squares technique returns no solution for the singular case and extremely large parameters for the close-to-singular case.

CB Predictor uses the matrix technique for singular value decomposition. Starting with:

$$\mathbf{y} = \mathbf{X} * \mathbf{b}$$

This can be rewritten as:

$$y = \begin{bmatrix} U \end{bmatrix} \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & w_n \end{bmatrix} \begin{bmatrix} V^T \end{bmatrix} \bullet b$$

Solving this, the regression coefficients are:

$$b = \begin{bmatrix} V \end{bmatrix} \begin{bmatrix} \frac{1}{w_j} & 0 & 0 & 0 \\ 0 & \frac{1}{w_j} & 0 & 0 \\ 0 & 0 & \frac{1}{w_j} & 0 \\ 0 & 0 & 0 & \frac{1}{w_j} \end{bmatrix} \begin{bmatrix} U^T \end{bmatrix} \begin{bmatrix} y \end{bmatrix}$$



Appendix E

Regression Statistic Formulas

In this appendix

This appendix provides formulas for the various regression statistics that CB Predictor calculates:

- R^2
- Adjusted R^2
- SSE
- F
- t
- p

Regression statistics

The statistics used to analyze a regression are different from the statistics used to analyze a time-series forecast. The formulas for these regression statistics appear below.

R^2

R^2 is the coefficient of determination. This statistic represents the proportion of error for which the regression accounts.

There are many methods to calculate R^2 . CB Predictor uses the equation:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Adjusted R^2

You can calculate a regression equation by using the same number of data points as you have equation coefficients. However, the regression equation will not be as universal as a regression equation calculated using three times the number of data points as equation coefficients.

Glossary Term:
degrees of freedom—
The number of data
points minus the
number of estimated
parameters
(coefficients).

To correct the R^2 for such situations, an adjusted R^2 takes into account the degrees of freedom of an equation. When you suspect your R^2 is higher than it should be, calculate the R^2 and adjusted R^2 . If the R^2 and the adjusted R^2 for your equation are close, then your R^2 is probably accurate. If R^2 is much higher than the adjusted R^2 , then you probably don't have enough data points to calculate the regression accurately.

The formula for adjusted R^2 is:

$$\text{Adjusted } R^2 = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}$$

where n is the number of data points you have and k is the number of independent variables.

To calculate the adjusted R^2 for a set of data:

1. **Subtract one from the number of data points.**
2. **Calculate the number of data points minus the number of independent variables minus 1.**
3. **Divide the result from step 1 by the result from step 2.**
4. **Calculate R^2 for the data.**

For more information, see “ R^2 ” on page 146.

5. **Calculate 1 minus R^2 .**
6. **Multiply the result of step 5 by the result of step 3.**
7. **Calculate 1 minus the last result.**

SSE

The formula for SSE is:

$$\text{SSE} = \sum_{i=1}^n e_i^2$$

where n is the number of data points you have and e is the unexplained error (the deviation between the actual point and the forecasted point on the line) for each point.

To calculate the SSE for a set of data:

1. **Measure the total deviation between each point and the line.**
2. **Square the deviations**
3. **Add together the squared deviations for each.**

F

After finding a regression equation, the F statistic checks the significance of the relationship between the dependent variable and the particular combination of independent variables in the regression equation. The F statistic is based on the scale of the Y values, so analyze this statistic in combination with the P value (described in the next section). When comparing the F statistics for similar sets of data with the same scale, the higher F statistic is better.

The formula for the F statistic is:

$$F = \frac{\sum (\hat{Y}_i - \bar{Y})^2 / (m - 1)}{\sum (Y_i - \hat{Y}_i)^2 / (n - m)}$$

where Y_i is the actual value of a point for a given time period i , \bar{Y} is the mean of the data, n is the total number of fitted points, and \hat{Y}_i is the fitted forecast value for the time period i , m is the number of coefficients in the regression equation, and n is the number of data points.

To calculate the F statistic for a set of data:

1. Calculate the difference between each forecasted point (\hat{Y}_i) and the mean(\bar{Y}).
2. Square the difference.
3. Repeat steps 1 through 2 for each fitted point and sum the results.
4. Subtract 1 from the number of coefficients.
5. Divide the result of step 3 by the result of step 4.
6. Calculate the difference between each forecasted point and the actual point (Y_i).
7. Square the difference.
8. Repeat steps 6 and 7 for each fitted point and sum the results.
9. Subtract the number of number of coefficients from the number of data points.
10. Divide the result of step 8 by the result of step 9.
11. Divide the result of step 5 by the result of step 10.

t

If your F statistic and p indicate a significant relationship between the dependent and the independent variables as a whole, you then want to see the significance of the relationship of the dependent variable to each independent variable. The t statistic tests for the significance of the specified independent variable in the presence of the other independent variables.

The formula for the t statistic is:

$$t = \frac{b_p}{se(b_p)}$$

where b_p is the coefficient to check and $se(b_p)$ is the standard error of the coefficient. To calculate the standard error of the coefficient, CB Predictor uses the matrix technique. Start with the regression equation:

$$y = b_0 + b_1x_{t1} + b_2x_{t2} + b_3x_{t3} + \dots + b_px_{tp} +$$

e

where t is the time period and p is the index of the independent variable. Define the matrix:

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

The standard error of the coefficient is:

$$s_{b_p} = s\sqrt{c_{pp}}$$

where c_{pp} is a diagonal element of the matrix $(X'X)^{-1}$ and the standard error, s , for y_t is:

$$s = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n - (p + 1)}}$$

p

CB Predictor calculates p and displays it with the other statistics. However, you can look up p for any F statistic using a table of critical F statistic values, available in most forecasting books. When using a critical F statistic table, you need to know the F statistic and the explained and unexplained degrees of freedom ($m - 1$ and $n - m$, respectively).



Appendix F

Error Messages

In this appendix

This appendix lists, alphabetically by first word, messages that might be displayed when you are entering or selecting data or running simulations. Either an explanation, a solution, or both are provided for each message. Messages not listed in this appendix should be self-explanatory.

Error messages

There are several problems that you might find while running CB Predictor. Below is a list of all the error messages, in alphabetical order, the reason the error message appeared, and the steps to try to resolve the problem.

Table F.1 CB Predictor error messages

<i>Message</i>	<i>Reason</i>	<i>Solution</i>
Bad value for seasonality.	You entered an invalid value for the number of periods for each season.	In the Data tab, make sure that the seasonality is an integer between 1 and either 100 or half the number of historical data points for the series, whichever is smaller.
Cannot find any data in range.	You selected only blank cells or your selected range only includes data in the first row or column that you indicated was your date row or column.	Reselect your data. Or, on the Data tab, uncheck the setting to treat the first row or column as the dates.
Cannot find any data to forecast in the specified range. Reselect your data or uncheck the Treat First [Row/Column] As Date setting.	Your selected range doesn't include data.	Reselect your data.
Cannot perform MLR until you identify your regression variables (no independent or dependent variables).	To use regression, you must identify at least one dependent variables and one independent variable.	Open the Regression Variables dialog and select at least one dependent and one independent variable.
Could not allocate memory for calculating the Single Exponential Smoothing starting points. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Additive Seasonal No Trend method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Double Exponential Smoothing method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.

Table F.1 CB Predictor error messages

<i>Message</i>	<i>Reason</i>	<i>Solution</i>
Could not allocate memory for calculating the Double Moving Average method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Holt-Winters' Additive method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Holt-Winters' Multiplicative method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Multiplicative Seasonal No Trend method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Single Exponential Smoothing method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Could not allocate memory for calculating the Single Moving Average method. Reduce the number of historical data series.	You have run into a program limitation that will keep you from forecasting all your historical series at once.	Select fewer data series to forecast at a time.
Error establishing a connection with Excel. Close and restart your forecasting application.	Some problem corrupted the communication between CB Predictor and Excel.	Close the CB Predictor wizard and restart it. Or, close Excel and restart both Excel and CB Predictor. Or, close Windows and restart Windows, Excel, and CB Predictor.
Error in data range.	There is some error in your selected data.	Check your selected data range and eliminate unusual characters, values, or formula cells.

Table F.1 CB Predictor error messages

<i>Message</i>	<i>Reason</i>	<i>Solution</i>
Invalid value for lag.	You entered an invalid string in the Lag field. The lag must be a positive integer.	Enter a positive integer in the Lag field.
Lag is too large or too small for the amount of data selected.	The value in the Lag field is invalid for your data. The lag must be an integer greater than zero and less than one third the number of data points in the series.	Make sure the number in the Lag field is an integer greater than zero and less than one third the number of data points in the series.
Lead or holdout must be greater than 0 and less than 25% of the number of data points.	You entered an invalid number in the Simple Lead, Weighted Lead, or Holdout field.	Make sure you enter an integer greater than 0 and less than 25% of the number of data points in the Simple Lead, Weighted Lead, or Holdout field.
No active methods. Please select methods from the Method Gallery.	You don't have any methods selected from the Method Gallery.	In the Method Gallery, select at least one method to use to forecast your data.
One test must be chosen. Choose either R-squared or F-test or both.	You must select at least one stopping criteria for stepwise regression.	Select either the R^2 or F-test stopping criteria.
Probability to remove must be at least 0.05 greater than the probability to add.	The value in the Probability To Remove field must be at least 0.05 greater than the value in the Probability To Add field.	Make sure the value in the Probability To Remove field is at least 0.05 more than the value in the Probability To Add field.
Selected value used to determine best fit is not MAD, MAPE, or RMSE. Select one of these.	This indicates that there is a problem with fit selection method.	In the Method Gallery tab, click the Fit button and make sure that one of the error fitting techniques in the dialog is selected. If it still doesn't work, call Decisioneering technical support.
Some values in the data range are too high or too low.	You cannot use values above 1×10^{20} or below 1×10^{-20} .	Remove values above 1×10^{20} or below 1×10^{-20} . If these values are necessary for your calculations, try scaling your data.

Table F.1 CB Predictor error messages

<i>Message</i>	<i>Reason</i>	<i>Solution</i>
The forecasting program had a problem finding the best parameters for Double Exponential Smoothing. Set the parameters manually.	Your historical data had some unusual pattern that kept CB Predictor from automatically calculating the optimal parameters for Double Exponential Smoothing.	Check your data and eliminate outliers. Or, try manually setting the Double Exponential Smoothing parameters. Or, uncheck Double Exponential Smoothing in the Method Gallery tab.
The forecasting program had a problem finding the best parameters for Single Exponential Smoothing. Set the parameters manually.	Your historical data had some unusual pattern that kept CB Predictor from automatically calculating the optimal parameters for Single Exponential Smoothing.	Check your data and eliminate outliers. Or, try manually setting the Single Exponential Smoothing parameters. Or, uncheck Single Exponential Smoothing in the Method Gallery tab.
The forecasting program had a problem finding the best parameters for the Additive Seasonal No Trend method. Set the parameters manually.	Your historical data had some unusual pattern that kept CB Predictor from automatically calculating the optimal parameters for Additive Seasonal No Trend.	Check your data and eliminate outliers. Or, try manually setting the Additive Seasonal No Trend parameters. Or, uncheck Additive Seasonal No Trend in the Method Gallery tab.
The forecasting program had a problem finding the best parameters for the Multiplicative Seasonal No Trend method. Set the parameters manually.	Your historical data had some unusual pattern that kept CB Predictor from automatically calculating the optimal parameters for Multiplicative Seasonal No Trend.	Check your data and eliminate outliers. If that doesn't help, try manually setting the Multiplicative Seasonal No Trend parameters. Or, uncheck Multiplicative Seasonal No Trend in the Method Gallery tab.
The method parameter must be between 0.001 and 0.999.	You entered an invalid value for one of the method parameters.	Make sure all the method parameter fields (alpha, beta, or gamma) have values between 0.001 and 0.999.

Table F.1 CB Predictor error messages

<i>Message</i>	<i>Reason</i>	<i>Solution</i>
The minimum number of data points for any forecast method is 5.	<p>You must have at least 5 historical data points to use time-series forecasting.</p> <p>Or, the number of points after excluding the holdout points is less than 5.</p>	<p>Make sure you have at least 5 historical data points.</p> <p>If you have more than 5 data points, check the Forecasting Technique in the Advanced Options dialog. If you are using Holdout, lower the number of holdout points or use a different technique.</p>
The number of periods to forecast must be an integer between 1 and 100.	You entered an invalid number for the number of periods to forecast. Your number might be negative, a decimal number, or less than one.	On the Results tab, change the number of forecasted data points to an integer of at least 1.
The paste range value entered is not valid.	In Step 9, the Paste Forecasts At Cell field requires a cell or range of cells.	Enter a cell (such as F6) or range of cells (such as H1:H5).
The size must be an integer between 20% and 400%.	When specifying the size of a chart, you must use an integer between 20 and 400 to represent the percentage.	In the Preferences dialogs, make sure the Chart Size field has an integer between 20 and 400.
There is at least one non-numeric cell in the data range.	Your data, not including headers and dates, have at least one cell with illegal characters in it.	<p>If you have headers and dates, make sure that the header and date settings on the Data tab are checked.</p> <p>Or, check your historical data, not including headers and dates, for cells with formulas or characters instead of numbers.</p>
Unrecognized error.	You encountered an undocumented error.	Call Decisioneering technical support and describe how you found the error.
Value must be between zero and one.	The Threshold and all the Probability values must be between 0.001 and 0.999.	Make sure the Threshold, Probability, Probability To Add, and Probability To Remove fields all have values between 0.001 and 0.999.

Table F.1 CB Predictor error messages

<i>Message</i>	<i>Reason</i>	<i>Solution</i>
You cannot use Double Moving Average to forecast more than 33% of the original number of historical data points. Lower the Periods parameter (by double-clicking the DMA icon in the Method Gallery) to less than 33% of the number of historical data points.	You are trying to forecast more data points than is reasonable for the Double Moving Average method with your number of historical data points.	In the DMA Parameters dialog, lower the Periods parameter to less than 33% of the number of historical data points. Or, use more historical data. Or, on the Method Gallery tab, unselect Double Moving Average.
You cannot use Single Moving Average to forecast more than 50% of the original number of historical data points. Lower the Periods parameter (by double-clicking the SMA icon in the Method Gallery) to less than 50% of the number of historical data points.	You are trying to forecast more data points than is reasonable for the Single Moving Average method with your number of historical data points.	In the SMA Parameters dialog, lower the Periods parameter to less than 50% of the number of historical data points. Or, use more historical data. Or, on the Method Gallery tab, unselect Single Moving Average.
You must use at least 6 historical data periods to use Double Moving Average. Use more historical data or unselect Double Moving Average in the Method Gallery.	Double Moving Average requires at least 6 historical data points.	Use at least 6 historical data points. Or, uncheck Double Moving Average in the Method Gallery tab.



Bibliography

In this bibliography

Bibliography entries by subject:

- Forecasting
- Regression analysis

Forecasting

Bowerman, B.L., and R.T. O'Connell (contributor). *Forecasting and Time Series: An Applied Approach* (The Duxbury Advanced Series in Statistics and Decision Sciences). Belmont, CA: Duxbury Press, 1993.

DeLurgio, S.A., Sr. *Forecasting Principles and Applications*. Boston: Irwin/McGraw-Hill, 1998.

Makridakis, S., S.C. Wheelwright, and R.J. Hyndman. *Forecasting Methods and Applications*. 3rd ed. New York: John Wiley & Sons, 1998.

Montgomery, D.C., and L.A. Johnson. *Forecasting and Time Series Analysis*. New York: McGraw-Hill, 1976.

Regression analysis

Dongarra, J.J., et al. *LINPACK User's Guide*. Philadelphia: S.I.A.M., 1979.

Forsythe, G.E., M.A. Malcolm, and C.B. Moler. *Computer Methods for Mathematical Computations*. Englewood Cliffs, NJ: Prentice-Hall, 1977.

Golub, G.H., and C.F. Van Loan. *Matrix Computations*, 2nd ed. Baltimore: Johns Hopkins University Press, 1989.

Ott, L. *An Introduction to Statistical Methods and Data Analysis*. Boston: PWS-KENT, 1988.

Smith, B.T., et al. *Matrix Eigensystem Routines - EISPACK Guide*, 2nd ed., vol. 6 of Lecture Notes in Computer Science. New York: Springer-Verlag, 1976.

Wilkinson, J.H., and C. Reinsch. "Linear Algebra," vol. II of *Handbook for Automatic Computation*. New York: Springer-Verlag, 1971.



Glossary

In this glossary

A compilation of terms specific to CB Predictor and Crystal Ball as well as statistical terms used in this manual.

Adjusted R^2

Corrects R^2 to account for the degrees of freedom in the data.

assumptions

Estimated values in a spreadsheet model that Crystal Ball defines with a probability distribution.

autocorrelation

Describes a relationship or correlation between values of the same data series at different time periods.

autoregression

Describes a relationship similar to autocorrelation, except instead of the variable being related to other independent variables, it is related to previous values of its own data series.

causal methods

A relationship between two variables where changes in one independent variable not only correspond to a particular increase or decrease in the dependent variable, but actually cause the increase or decrease.

Crystal Ball

Decisioneering's risk analysis add-in for Excel.

Crystal Ball forecast

A statistical summary of the assumptions in a spreadsheet model, output graphically or numerically.

degrees of freedom

The number of data points minus the number of estimated parameters (coefficients).

dependent variable

In multiple linear regression, a data series or variable that depends on another data series. You should use multiple linear regression as the forecasting method for any dependent variable.

DES

Double exponential smoothing.

double exponential smoothing

Double exponential smoothing applies single exponential smoothing twice, once to the original data and then to the resulting single exponential smoothing data. It is useful where the historic data series is not stationary.

double moving average

Smooths out past data by performing a moving average on a subset of data that represents a moving average of an original set of data.

Durbin-Watson

Tests for autocorrelation of one time lag.

error

The difference between the actual data values and the forecasted data values.

forecast

The prediction of values of a variables based on known past values of that variable or other related variables. Forecasts can also describe predicted values based on Crystal Ball spreadsheet models and expert judgements.

forward stepwise

A regression method that adds one independent variable at a time to the multiple linear regression equation, starting with the independent variable with the greatest significance.

***F*-test statistic**

See *F* statistic.

***F* statistic**

Tests the overall significance of the multiple linear regression equation.

holdout

Optimizes the forecasting parameters to minimize the error measure between a set of excluded data and the forecasting values. CB Predictor doesn't use the excluded data to calculate the forecasting parameters.

Holt-Winters' additive seasonal smoothing

Separates a series into its component parts: seasonality, trend and cycle, and error. This method determines the value of each, projects them forward, and reassembles them to create a forecast.

Holt's double exponential smoothing

Similar to double exponential smoothing, but this method lets you use a different smoothing constant for the second smoothing process.

Holt-Winters' multiplicative seasonal smoothing

Considers the effects of seasonality to be multiplicative, that is, growing (or decreasing) over time. This method is similar to the Holt-Winters' additive method.

HyperCasting

CB Predictor's ability to create a regression equation, forecast independent variables, and combine the forecasts and equation to forecast a dependent variable in one step.

hyperplane

A geometric plane that spans more than two dimensions.

independent variable

In multiple linear regression, the data series or variables that influence the another data series or variable.

Intelligent Input

CB Predictor's ability to select all the adjacent data, headers, and dates to define the selection to forecast. The Intelligent Input also determines whether the data is in rows or columns and whether there are dates or headers. To use the Intelligent Input, select one cell in your data range before starting the CB Predictor wizard.

iterative stepwise regression

A regression method that adds or subtracts one independent variable at a time to or from the multiple linear regression equation.

lag

Defines the offset when comparing a data series with itself. For autocorrelation, this refers to the offset of data that you choose when correlating a data series with itself.

lead

A type of forecasting that optimizes the forecasting parameters to minimize the error measure between the historical data and the fit values, offset by a specified number of periods (lead).

least-squares approach

Measures how closely a line matches a set of data. This approach measures the distance of each actual data point from the line, squares each distance, and adds up the squares. The line with the smallest square deviation is the closest fit.

level

A starting point for the forecast. For a set of data with no trend, this is equivalent to the y -intercept.

linear equation

An equation with only linear terms. A linear equation has no terms containing variables with exponents or variables multiplied by each other.

linear regression

A process that models a variable as a function of other first-order explanatory variables. In other words, it approximates the curve with a line, not a curve, which would require higher-order terms involving squares and cubes.

MAD

Mean absolute deviation. This is an error statistic that average distance between each pair of actual and fitted data points.

MAPE

Mean absolute percentage error. This is a relative error measure that uses absolute values to keep the positive and negative errors from cancelling out each other and uses relative errors to let you compare forecast accuracy between time-series models.

multiple linear regression

A case of linear regression where one dependent variable is described as a linear function of more than one independent variable.

naive forecast

A forecast obtained with minimal effort based on only the most recent data; e.g., using the last data point to forecast the next period.

 p

Indicates the probability of obtaining an F or t statistic as large as the one calculated for your data.

partial F statistic

Tests the significance of a particular independent variable within the existing multiple linear regression equation.

PivotTable

An interactive table in Excel. You can move rows and columns and filter PivotTable data.

regression

A process that models a dependent variable as a function of other explanatory (independent) variables.

residuals

The difference between the actual data and the predicted data for the dependent variable in multiple linear regression.

RMSE

Root mean squared error. This is an absolute error measure that squares the deviations to keep the positive and negative deviations from cancelling out each other. This measure also tends to exaggerate large errors, which can help when comparing methods.

 R^2

Coefficient of determination. This statistic indicates what proportion of the dependent variable error that the regression line explains.

seasonality

The change that seasonal factors cause in a data series. For example, if sales increase during the Christmas season and during the summer, the data is seasonal with a six-month period.

SES

Single exponential smoothing.

single exponential smoothing

Weights past data with exponentially decreasing weights going into the past; that is, the more recent the data value, the greater its weight. This largely overcomes the limitations of moving averages or percentage change models.

single moving average

Smooths out past data by averaging the last several periods and projecting that view forward. CB Predictor automatically calculates the optimal number of periods to be averaged.

singular value decomposition

A method that solves a set of equations for the coefficients of a regression equation.

smoothing

Estimates a smooth trend by removing extreme data and reducing data randomness.

SSE

Sum of square deviations. The least squares technique for estimating regression coefficients uses this statistic, which measures the error not eliminated by the regression line.

SVD

Singular value decomposition.

time series

A set of values that are ordered in equally spaced intervals of time.

trend

A long-term increase or decrease in time-series data.

 t statistic

Tests the significance of the relationship between the dependent variable and any individual independent variable, in the presence of the other independent variables.

variables

In regression, data series are also called variables.

weighted lead

A type of forecasting that optimizes the forecasting parameters to minimize the average error measure between the historical data and the fit values, offset by several different periods (leads).



Index

In this index

A comprehensive index designed to give you quick access to the information in this manual.

Index

A

- additive methods
 - Holt-Winters' seasonal smoothing 40
 - seasonal smoothing 39
- adjusted R squared
 - description 50
 - formula 146
- Advanced button 111
- Advanced Options dialog 113
- All Series list 107
- alpha
 - for nonseasonal methods 38
 - for seasonal smoothing methods 41
- assumptions
 - description 94
 - generating 94
- asterisk in Method Gallery 69
- autocorrelation
 - formula 139
 - statistics 138
- autocorrelation in regression 51
- Autocorrelations view 103
- Automatic option, for parameters 112

B

- bakery example 82
- best methods, determining error measures 69
- beta
 - for Holt's double moving average 38
 - for seasonal smoothing methods 41

C

- captures, screen 5
- CB Predictor
 - 10-step process 58
 - Intelligent Input 60
 - starting 57
 - with Crystal Ball 94
- cells, handling empty 59
- charts 75
 - customizing 79
 - preferences 121
- Clear All button 111
- coefficients
 - finding regression 47, 142
 - of determination 50
 - testing coefficients 51
- confidence interval
 - formula 136

- confidence interval, setting 71
- consulting service 4
- conventions, manual 4
- creating spreadsheets 56
- criteria, setting stopping 109
- criteria, stopping stepwise 48
- Crystal Ball 94
 - assumptions 94
 - example 96
 - forecasts 94
- customizing reports and charts 79

D

- data
 - arrangement 61
 - historical statistics 52
 - minimum historical 59
 - selecting historical 59
 - viewing historical 61
 - with empty cells 59
- Data Attributes tab 104
 - step 4 63
 - step 5 64
- Data In Rows/Columns option 101
- dates
 - formatting 56
 - including 101
- decomposition, singular value 47
- dependent variables 64
- DES, *see* Holt's double exponential smoothing
- deviation, standard 52
- dialogs
 - Stepwise Options 108
 - View Historical Data 102
- DMA, *see* double moving average
- Do Not Forecast option 108
- double exponential smoothing, *see* Holt's double exponential smoothing
- double moving average 36
- Durbin-Watson
 - description 51
 - formula 138

E

- equations
 - adjusted R squared 146
 - Durbin-Watson 138
 - F statistic 147
 - Holt's double exponential smoothing 128

- Holt-Winters' additive seasonal smoothing 130
- Holt-Winters' multiplicative seasonal smoothing 131
- least squares 142
- linear regression 46
- MAD 135
- MAPE 135
- nonseasonal methods 128
- R squared 146
- regression statistics 146
- RMSE 134
- seasonal additive 129
- seasonal methods 129
- seasonal multiplicative 129
- single exponential smoothing 128
- singular value decomposition 143
- SSE 147
- t statistic 149
- Theil's U 137
- error measures
 - MAD 43
 - MAD formula 135
 - MAPE 43
 - MAPE formula 135
 - RMSE 42
 - RMSE formula 134
 - selecting 69
- errors
 - messages 152
 - time-series measures 42
- example file names 5
- examples
 - financial 86
 - human resources 88
 - inventory control 82
 - Monica's bakery 82
 - shampoo sales 9
 - Toledo gas 17
 - using Crystal Ball 96
- F**
 - F statistic
 - as stopping criteria
 - description 51
 - formula 147
 - financial example 86
 - First Column (Row) Has Headers 101
 - First Row (Column) Has Dates 101
 - forecasting
 - categories of 34
 - holdout 45
 - maximum number of series 56
 - references 160
 - simple lead 44
 - techniques 43
 - time-series 8, 35
 - weighted lead 44
 - forecasting methods
 - nonseasonal formulas 128
 - seasonal formulas 129
 - forecasting statistics
 - Theil's U 46
 - forecasting techniques
 - holdout 45
 - simple lead 44
 - standard 43
 - weighted lead 44
 - forecasts
 - Crystal Ball 94
 - previewing and running 73
 - formulas
 - adjusted R squared 146
 - Durbin-Watson 138
 - F statistic 147
 - Holt's double exponential smoothing 128
 - Holt-Winters' additive seasonal smoothing 130
 - Holt-Winters' multiplicative seasonal smoothing 131
 - least squares 142
 - MAD 135
 - MAPE 135
 - nonseasonal methods 128
 - R squared 146
 - regression statistics 146
 - RMSE 134
 - seasonal additive smoothing 129
 - seasonal methods 129
 - seasonal multiplicative smoothing 129
 - single exponential smoothing 128
 - singular value decomposition 143
 - SSE 147
 - t statistic 149
 - Theil's U 137
 - time-series methods 128
 - forward stepwise regression 48
 - F-test, *see F statistic*

Index

G

gamma, for seasonal smoothing methods 41
gas tutorial 17

H-K

hardware requirements 2
headers 101
historical data
 arrangement 61
 minimum required 59
 selecting 59
 statistics 52
 viewing 61
holdout forecasting 45
Holt's double exponential smoothing
 description 37
 formulas 128
Holt-Winters' additive seasonal smoothing
 description 40
 formulas 130
Holt-Winters' multiplicative seasonal smoothing
 description 41
 formulas 131
how the program works 16
how this manual is organized 2
human resources example 88
HyperCasting 64
hyperplanes 47
identifying time periods and seasonality 63
independent variables
 defining 64
 setting lag 107
Input Data tab 100
 step 1 59
 step 2 61
 step 3 61
Input, Intelligent 60
Intelligent Input 60
intervals, confidence 71
inventory example 82
iterative stepwise regression 49

L

lag option 107
lead forecasting
 simple 44
 weighted 44
least squares 142
linear regression

 description 46
 steps 19, 64
linear smoothing 36
Ljung-Box statistic 53
 description 53
 formula 139

M

MAD
 description 43
 formula 135
manual
 conventions 4
 organization 2
MAPE
 description 43
 formula 135
maximum 53
maximum number of forecasts 56
mean 52
mean absolute deviation, *see* MAD
mean absolute percentage error, *see* MAPE
measures
 selecting error 69
 time-series error 42
measures, error
 MAD 43
 MAPE 43
 RMSE 42
messages, error 152
Method Gallery 110
methods
 nonseasonal formulas 128
 regression 48
 seasonal 67
 seasonal additive 39
 seasonal formulas 129
 seasonal multiplicative 39
 seasonal smoothing 39
 selecting time-series 66
 setting regression 106
 standard regression 48
methods tables
 description 77
 preferences 123
minimum 53
minimum historical data 59
moving averages
 double 36

- periods 38
 - single 36
- multiple linear regression
 - description 46
 - steps 19, 64
- multiplicative methods
 - Holt-Winters' seasonal smoothing 41
 - seasonal smoothing 39

N

Number Of Periods To Forecast field 115

O

- options
 - Do Not Forecast 108
 - lag 107
- Override Best button 126

P

- parameters
 - automatic 112
 - dialogs 111
 - nonseasonal methods 38
 - seasonal smoothing 41
 - user-defined 112
- pasted forecasts 74
 - preferences 118
 - values as assumptions 94
- periods
 - moving average parameter 38
 - number to forecast 71
- PivotTables
 - methods 77
 - results 76
- Preferences
 - button 116
 - dialog 117
- preferences
 - charts 121
 - methods table 123
 - paste 118
 - reports 119
 - results table 122
- Preview
 - button 102
 - dialog 125
- previewing forecasts 73
- probability 51

R

- R squared
 - as stopping criteria 48
 - description 50
 - formula 146
- R squared, adjusted 50
- Range field 101
- ranges with empty cells 59
- Raw Data view 102
- references 160
- referral service 4
- regression 46
 - adjusted R squared 50
 - detecting autocorrelation 51
 - determining coefficients 47
 - Durbin-Watson 51
 - equation 46
 - F statistic 51
 - finding coefficients 142
 - forward stepwise 48
 - methods 48
 - p 51
 - pasted assumptions 95
 - R squared 50
 - SSE 50
 - statistics 50
 - steps 19, 64
 - t statistic 51
 - testing equation 51
 - testing probability 51
 - using 64
- regression methods
 - iterative stepwise 49
 - standard 48
- regression methods, setting 106
- Regression Variables dialog 106
- reports 76
 - customizing 79
 - preferences 119
- requirements, system 2
- results
 - charts 75
 - methods table 77
 - pasted forecasts 74
 - reports 76
 - results table 76
 - selecting 72
 - title 116
- Results tab 115

Index

- step 10 73
- step 7 71
- step 8 71
- step 9 72
- results tables
 - description 76
 - preferences 122
- RMSE
 - description 42
 - formula 134
- root mean squared error, *see* RMSE
- Run button 102
- running forecasts 73
- S**
- screen captures 5
- seasonal additive smoothing
 - description 39
 - formulas 129
- seasonal methods 67
 - Holt-Winters' additive seasonal smoothing 40
 - Holt-Winters' multiplicative seasonal smoothing 41
 - seasonal additive smoothing 39
 - seasonal multiplicative smoothing 39
- seasonal multiplicative smoothing
 - description 39
 - formulas 129
- seasonal smoothing
 - methods 39
 - parameters 41
- seasonality, identifying 63
- Select All button 111
- Select button
 - on Input Data tab 101
 - on Results tab 116
- Select Variables button 106
- selecting
 - historical data 59
 - methods 66
 - results 72
- services
 - consulting referral 4
 - technical support 3
- SES, *see* single exponential smoothing
- setting method parameters 112
- setting stopping criteria 109
- shampoo tutorial 9
- simple lead forecasting 44
- single exponential smoothing
 - description 37
 - formula 128
- single moving average 36
- singular value decomposition
 - formulas 143
 - use 47
- SMA, *see* single moving average
- smoothing
 - linear 36
 - seasonal 39
- smoothing methods
 - Holt's double exponential 37
 - Holt-Winters' additive seasonal 40
 - Holt-Winters' multiplicative seasonal 41
 - seasonal 39
 - seasonal additive 39
 - seasonal multiplicative 39
 - single exponential 37
- software requirements 2
- spreadsheets
 - creating 56
 - titles 56
- SSE
 - description 50
 - formula 147
- standard
 - forecasting technique 43
 - regression 48
- standard deviation 52
- starting CB Predictor 57
- statistics
 - adjusted R squared 50
 - Durbin-Watson 51
 - Durbin-Watson formula 138
 - F 51
 - historical data 52
 - Ljung-Box 53, 139
 - maximum 53
 - mean 52
 - minimum 53
 - p 51
 - R squared 50
 - R squared formula 146
 - regression 50
 - regression formulas 146
 - SSE 50
 - SSE formula 147

- standard deviation 52
- t 51
- t formula 149
- Theil's U 46
- Theil's U formula 137
- viewing for historical data 101
- steps
 - 1 59
 - 10 73
 - 2 61
 - 3 61
 - 4 63
 - 5 64
 - 6 66
 - 7 71
 - 8 71
 - 9 72
- process of 10 58
- Stepwise Options dialog 108
- stepwise regression
 - forward 48
 - iterative 49
 - stopping criteria 48
- stopping criteria 48
- setting 109
- sum of square deviations 50
- support, technical 3
- system requirements 2

T

- t statistic
 - description 51
 - formula 149
- tables
 - methods 77
 - results 76
- technical support 3
- techniques
 - forecasting 43
- templates, report and chart 79
- Theil's U
 - description 46

- formula 137
- time periods
 - identifying 63
 - number to forecast 71
- time-series
 - error measures 42
 - forecasting overview 8
- time-series forecasting, *see* forecasting
- time-series methods
 - pasted assumptions 95
 - selecting 66
 - specialties 67
- titles
 - results 116
 - spreadsheet 56
- Toledo Gas tutorial 17
- training service 3
- tutorials
 - shampoo sales 9
 - Toledo gas 17

U

- user-defined parameters 112

V

- variables
 - dependent and independent 64
 - regression 64
 - setting independent lag 107
- View Data button 101
- View Historical Data dialog 102
- viewing historical data 61

W

- weighted lead forecasting 44
- what you need 2
- who should use this program 1
- who this program is for 1
- worksheets, creating 56

Z

- zeros in data 59



Credits

CB Predictor User Manual

Written by Carol Werckman and Carl Crosswhite. Edited by Cameron Harris and Barbara Gentry.

Illustrations and screen captures by Barbara Gentry using Jasc, Inc. Paint Shop Pro.

This document was created electronically using Adobe Systems Incorporated Framemaker Release 7.1 for Microsoft Windows.

Typeset using NewBskvll BT and Univers fonts.

